Statistical analysis on testing of an entangled state based on Poisson distribution framework

Masahito Hayashi*,†, Akihisa Tomita*,‡, and Keiji Matsumoto§,*

*ERATO-SORST Quantum Computation and Information Project,
Japan Science and Technology Agency (JST), Tokyo 113-0033, Japan
†Superrobust Computation Project, Information Science and Technology Strategic Core (21st Century COE by MEXT)
Graduate School of Information Science and Technology, The University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan
‡Fundamental Research Laboratories, NEC, Tsukuba 305-8501, Japan
§National Institute of Informatics, Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

A hypothesis testing scheme for entanglement has been formulated based on the Poisson distribution framework instead of the POVM framework. Three designs were proposed to test the entangled states in this framework. The designs were evaluated in terms of the asymptotic variance. It has been shown that the optimal time allocation between the coincidence and anti-coincidence measurement bases improves the conventional testing method. The test can be further improved by optimizing the time allocation between the anti-coincidence bases.

I. INTRODUCTION

Entangled states are an essential resource for various quantum information processings[1, 2]. Hence, it is required to generate maximally entangled states. However, for a practical use, it is more essential to guarantee the quality of generated entangled states. Statistical hypothesis testing is a standard method for guaranteeing the quality of industrial products. Therefore, it is much needed to establish the method for statistical testing of maximally entangled states.

Quantum state estimation and quantum state tomography are known as the method of identifying the unknown state[3, 4, 5]. Quantum state tomography [6] has been recently applied to obtain full information of the 4×4 density matrix. However, if the purpose is testing of entanglement, it is more economical to concentrate on checking the degree of entanglement. Such a study has been done by Tsuda et al [8] as optimization problems of POVM. However, an implemented quantum measurement cannot be regarded as an application of a POVM to a single particle system or a multiple application of a POVM to single particle systems. In particular, in quantum optics, the following measurement is often realized, which is not described by a POVM on a single particle system. The number of generated particles is probabilistic. We prepare a filter corresponding to a projection P, and detect the number of particle passing through the filter. If the number of generated particles obeys a Poisson distribution, as is mentioned in Section II, the number of detected particles obeys another Poisson distribution whose average is given by the density and the projection P.

In this kind of measurements, if any particle is not detected, we cannot decide whether a particle is not generated or it is generated but does not pass through the filter. If we can detect the number of generated particles

as well as the number of passing particles, the measurement can be regarded as the multiple application of the POVM $\{P, I - P\}$. In this case, the number of applications of the POVM is the variable corresponding to the number of generated particles. Also, we only can detect the empirical distribution. Hence, our obtained information almost discuss by use of the POVM $\{P, I - P\}$.

However, if it is impossible to distinguish the two events by some imperfections, it is impossible to reduce the analysis of our obtained information to the analysis of POVMs. Hence, it is needed to analyze the performance of the estimation and/or the hypothesis testing based on the Poisson distribution describing the number of detected particles. If we discuss the ultimate bound of the accuracy of the estimation and/or the hypothesis testing, we do not have to treat such imperfect measurements. Since several realistic measurements have such imperfections, it is very important to optimize our measurement among such a class of imperfect measurements.

In this paper, our measurement is restricted to the detection of the number of the particle passing through the filter corresponding to a projection P. We apply this formulation to the testing of maximally entangled states on two qubit systems (two-level systems), each of which is spanned by two vectors $|H\rangle$ and $|V\rangle$. Since the target system is a bipartite system, it is natural to restrict to our measurement to local operations and classical communications (LOCC). In this paper, for a simple realization, we restrict our measurements to the number of the simultaneous detections at the both parties of the particles passing through the respective filters. We also restrict the total measurement time t, and optimize the allocation of the time for each filters at the both parties.

As our results, we obtain the following characterizations. If the average number of the generated particles is known, our choice is counting the coincidence events or the anti-coincidence events. When the true state is close to the target maximally entangled state

$|\Phi^{(+)}\rangle := \frac{1}{\sqrt{2}}(|HH\rangle + |VV\rangle)$ (that is, the fidelity between these is greater than $1/4$), the detection of anti-coincidence events is better than that of coincidence events. This result implies that the indistinguishability between the coincidence events and the non-generation event loses less information than that between the anti-coincidence events and the non-generation event.

This fact also holds even if we treat this problem taking into account the effect of dark counts. In this discussion, in order to remove the bias concerning the direction of the difference, we assume the equal time allocation among the vectors $\{|HV\rangle, |VH\rangle, |DX\rangle, |XD\rangle, |RR\rangle, |LL\rangle\}$, which corresponds to the anti-coincidence events, and that among the vectors $\{|HH\rangle, |VV\rangle, |DD\rangle, |XX\rangle, |RL\rangle, |LR\rangle\}$, which corresponds to the coincidence events, where $|D\rangle := \frac{1}{2}(|H\rangle + |V\rangle)$, $|X\rangle := \frac{1}{2}(|H\rangle - |V\rangle)$, $|R\rangle := \frac{1}{2}(|H\rangle + i|V\rangle)$, $|L\rangle := \frac{1}{2}(|H\rangle - i|V\rangle)$. Indeed, Barbieri et al [7] proposed to detect the anti-coincidence events for measuring an entanglement witness, they did not prove the superiority of detecting the anti-coincidence events in the framework of mathematical statistics.

However, the average number of the generated particles is usually unknown. In this case, we cannot estimate how close the true state is to the target maximally entangled state from the detection of anti-coincidence events. Hence, we need to count the coincidence events as additional information. in order to resolve this problem, we usually use the equal allocation between anti-coincidence events and coincidence events in the visibility method, which is a conventional method for checking the entanglement. However, since we measure the coincidence events and the anti-coincidence events based on one or two bases in this method, there is a bias concerning the direction of the difference. In order to remove this bias, we consider the detecting method with the equal time allocation among all vectors $\{|HV\rangle, |VH\rangle, |DX\rangle, |XD\rangle, |RR\rangle, |LL\rangle\}$ and $\{|HH\rangle, |VV\rangle, |DD\rangle, |XX\rangle, |RL\rangle, |LR\rangle\}$, and call it the modified visibility method.

In this paper, we also examine the detection of the total flux, which can be realized by detecting the particle without the filter. We optimize the time allocation among these three detections. We found that the optimal time allocation depends on the fidelity between the true state and the target maximally entangled state. If our purpose is estimating the fidelity F, we cannot directly apply the optimal time allocation. However, the purpose is testing whether the fidelity F is greater than the given threshold F_0, the optimal allocation at F_0 gives the optimal testing method.

If the fidelity F is less than a critical value, the optimal allocation is given by the allocation between the anti-coincidence vectors and the coincidence vectors (the ratio depends on F.) Otherwise, it is given by the allocation only between the anti-coincidence vectors and the total flux. This fact is valid even if the dark count exists. If the dark count is greater than a certain value, the optimal time allocation is always given by the allocation between the anti-coincidence vectors and the coincidence vectors.

Further, we consider the optimal allocation among anti-coincidence vectors when the average number of generated particles. The optimal allocation depends on the direction of the difference between the true state and the target state. Since the direction is usually unknown, this optimal allocation dose not seems useful. However, by adaptively deciding the optimal time allocation, we can apply the optimal time allocation. We propose to apply this optimal allocation by use of the two-stage method. Further, taking into account the complexity of testing methods and the dark counts, we give a testing procedure of entanglement based on the two-stage method. In addition, proposed designs of experiments were demonstrated by Hayashi et al. [14] in two photon pairs generated by spontaneous parametric down conversion (SPDC).

In this article, we reformulate the hypothesis testing to be applicable to the Poisson distribution framework, and demonstrate the effectiveness of the optimized time allocation in the entanglement test. The construction of this article is following. Section II defines the Poisson distribution framework and gives the hypothesis scheme for the entanglement. Section III gives the mathematical formulation concerning statistical hypothesis testing. Sections IV and V give the fundamental properties of the hypothesis testing: section IV introduces the likelihood ratio test and its modification, and section V gives the asymptotic theory of the hypothesis testing. Sections VI-IX are devoted to the designs of the time allocation between the coincidence and anti-coincidence bases: section VI defines the modified visibility method, section VII optimize the time allocation, when the total photon flux λ is unknown, section VIII gives the results with known λ, and section IX compares the designs in terms of the asymptotic variance. Section X gives further improvement by optimizing the time allocation between the anti-coincidence bases. Appendices give the detail of the proofs used in the optimization.

II. HYPOTHESIS TESTING SCHEME FOR ENTANGLEMENT IN POISSON DISTRIBUTION FRAMEWORK

Let $\mathcal{H}$ be the Hilbert space of our interest, and P be the projection corresponding to our filter. If we assume generation process on each time to be identical but individual, the total number n of generated particles during the time t obeys the Poisson distribution $\mathrm{Poi}(\lambda t)(n) := e^{-\lambda t}\frac{(\lambda t)^n}{n!}$. Hence, when the density of the true state is σ, the probability of the number k of de-

tected particles is given as

$$\sum_{n=0}^{\infty} \text{Poi}(\lambda t)(n) \binom{n}{k} (\mathbf{Tr}\, P\sigma)^k (1 - \mathbf{Tr}\, P\sigma)^{n-k}$$
$$= \text{Poi}(\lambda t\, \mathbf{Tr}\, P\sigma)(k). \tag{1}$$

$$\text{Poi}(\lambda t) \qquad P$$
$$\sigma$$
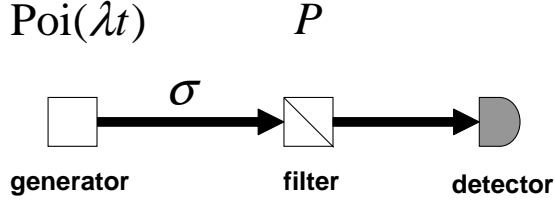
generator   filter   detector

FIG. 1: Experimental scheme in Poisson distribution framework

In fact, if we treat the Fock space generated by $\mathcal{H}$ instead of the single particle system $\mathcal{H}$, this measurement can be described by a POVM. However, since this POVM dooes not have a simple form, it is suitable to treat this measurement in the form (1).

Further, if we errorly detect the k' particles with the probability $\text{Poi}(\delta t)(k')$, the probability of the number k of detected particles is equal to

$$\sum_{k'=0}^{k} \text{Poi}(\lambda t\, \mathbf{Tr}\, P\sigma)(k - k') + \text{Poi}(\delta t)(k')$$
$$= \text{Poi}((\lambda\, \mathbf{Tr}\, P\sigma + \delta)t)(k).$$

This kind of incorrect detection is called dark count. Further, since we consider the bipartite case, i.e., the case where $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$, we assume that our projection P has the separable form $P_1 \otimes P_2$.

In this paper, under the above assumption, we discuss the hypothesis testing when the target state is the maximally entangled $|\Phi^{(+)}\rangle$ state while Usami et al.[11] discussed the state estimation under this assumption. Here we measure the degree of entanglement by the fidelity between the generated state and the target state:

$$F = \langle \Phi^{(+)} | \sigma | \Phi^{(+)} \rangle. \tag{2}$$

The purpose of the test is to guarantee that the state is sufficiently close to the maximally entangled state with a certain significance. That is, we are required to disprove that the fidelity F is less than a threshold F_0 with a small error probability. In mathematical statistics, this situation is formulated as hypothesis testing; we introduce the null hypothesis H_0 that entanglement is not enough and the alternative H_1 that the entanglement is enough:

$$H_0 : F \leq F_0 \text{ v.s. } H_1 : F > F_0, \tag{3}$$

with a threshold F_0.

Visibility is an indicator of entanglement commonly used in the experiments, and is calculated as follows: first, A's measurement vector $|x_A\rangle$ is fixed, then the measurement $|x_A, y_B\rangle$ is performed by rotating B's measurement vector $|x_B\rangle$ to obtain the maximum and minimum number of the counts, n_{max} and n_{min}. We need to make the measurement with at least two bases of A in order to exclude the possibility of the classical correlation. We may choose the two bases $\{|H\rangle, |V\rangle\}$ and $\{|D\rangle, |X\rangle\}$ as $|x_A\rangle$, for example. Finally, the visibility is given by the ratio between $n_{max} - n_{min}$ and $n_{max} + n_{min}$ with the respective A's measurement basis $|x_A\rangle$. However, our decision will contain a bias, if we choose only two bases as A's measurement basis $|x_A\rangle$. Hence, we cannot estimate the fidelity between the target maximally entangled state and the given state in a statistically proper way from the visibility.

Since the equation

$$|HH\rangle\langle HH| + |VV\rangle\langle VV| + |DD\rangle\langle DD|$$
$$+ |XX\rangle\langle XX| + |RL\rangle\langle RL| + |LR\rangle\langle LR|$$
$$= 2|\Phi^{(+)}\rangle\langle\Phi^{(+)}| + I \tag{4}$$

holds, we can estimate the fidelity by measuring the sum of the counts of the following vectors: $|HH\rangle, |VV\rangle, |DD\rangle, |XX\rangle, |RL\rangle$, and $|LR\rangle$, when λ is known[7, 8]. This is because the sum $n_1 := n_{HH} + n_{VV} + n_{DD} + n_{XX} + n_{RL} + n_{LR}$ obeys the Poisson distribution with the expectation value $(\lambda\frac{1+2F}{6} + \delta)t_1$, where the measurement time for each vector is $\frac{t_1}{6}$. We call these vectors the coincidence vectors because these correspond to the coincidence events.

However, since the parameter λ is usually unknown, we need to perform another measurement on different vectors to obtain additional information. Since

$$|HV\rangle\langle HV| + |VH\rangle\langle VH| + |XD\rangle\langle XD|$$
$$+ |DX\rangle\langle DX| + |RR\rangle\langle RR| + |LL\rangle\langle LL|$$
$$= 2I - 2|\Phi^{(+)}\rangle\langle\Phi^{(+)}| \tag{5}$$

also holds, we can estimate the fidelity by measuring the sum of the counts of the following vectors: $|HV\rangle, |VH\rangle, |DX\rangle, |XD\rangle, |RR\rangle$, and $|LL\rangle$. The sum $n_2 := n_{HV} + n_{VH} + n_{DX} + n_{XD} + n_{RR} + n_{LL}$ obeys the Poisson distribution $\text{Poi}((\lambda\frac{2-2F}{6} + \delta)t_2)$, where the measurement time for each vector is $\frac{t_2}{6}$. Combining the two measurements, we can estimate the fidelity without the knowledge of λ. We call these vectors the anti-coincidence vectors because these correspond to the anti-coincidence events.

We can also consider different type of measurement on λ. If we prepare our device to detect all photons, i.e., the case where the projection is $I \otimes I$, the detected number n_3 obeys the distribution $\text{Poi}((\lambda + \delta)t_3)$ with the measurement time t_3. We will refer to it as the total flux measurement. In the following, we consider the best time allocation for estimation and test on the fidelity, by applying methods of mathematical statistics. We will

assume that λ is known or estimated from the detected number n_3.

III. HYPOTHESIS TESTING FOR PROBABILITY DISTRIBUTIONS

A. Formulation

In this section, we review the fundamental knowledge of hypothesis testing for probability distributions[9]. Suppose that a random variable X is distributed according to a probability measure P_θ identified by the unknown parameter θ. We also assume that the unknown parameter θ belongs to one of mutually disjoint sets Θ_0 and Θ_1. When we want to guarantee that the true parameter θ belongs to the set Θ_1 with a certain significance, we choose the null hypothesis H_0 and the alternative hypothesis H_1 as

$$H_0 : \theta \in \Theta_0 \text{ versus } H_1 : \theta \in \Theta_1. \tag{6}$$

Then, our decision method is described by a test, which is described as a function $\phi(x)$ taking values in $\{0, 1\}$; H_0 is rejected if 1 is observed, and H_0 is not rejected if 0 is observed. That is, we make our decision only when 1 is observed, and do not otherwise. This is because the purpose is accepting H_1 by rejecting H_0 with guaranteeing the quality of our decision, and is not rejecting H_1 nor accepting H_1. Therefore, we call the region $\{x|\phi(1) = 1\}$ the rejection region. The test ϕ can be defined by the rejection region. In fact, we choosed the hypothesis that the fidelity is less than the given threshold θ_0 as the null hypothesis H_0 in Section II. This formulation is natural because our purpose is guaranteeing that the fidelity is not less than the given threshold θ_0.

From theoretical viewpoint, we often consider randomized tests, in which we probabilistically make the decision for a given data. Such a test is given by a function ϕ mapping to the interval $[0, 1]$. When we observe the data x, H_0 is rejected with the probability $\phi(x)$. In the following, we treat randomized tests as well as deterministic tests.

In the statistical hypothesis testing, we minimize error probabilities of the test ϕ. There are two types of errors. The type one error is the case where H_0 is rejected though it is true. The type two error is the converse case, H_0 is accepted though it is false. Hence, the type one error probability is given $P_\theta(\phi)$ $(\theta \in \Theta_0)$, and the type two error probability is given $1 - P_{\theta'}(\phi)$ $(\theta' \in \Theta_1)$, where

$$P_\theta(\phi) = \int \phi(x) dP_\theta(x).$$

It is in general impossible to minimize both $P_\theta(\phi)$ and $1 - P_{\theta'}(\phi)$ simultaneously because of a trade-off relation between them. Since we make our decision with guaranteeing its quality only when 1 is observed, it is definitively required that the type one error probability $P_\theta(\phi)$ is less than a certain constant α. For this reason, we minimize

the type two error probability $1 - P_{\theta'}(\phi)$ under the condition $P_\theta(\phi) \leq \alpha$. The constant α in the condition is called the risk probability, which guarantees the quality of our decision. If the risk probability is large enough, our decision has less reliability. Under this constraint for the risk probability, we maximize the probability to reject the hypothesis H_0 when the true parameter is $\theta' \in \Theta_1$. This probability is given as $P_\theta(\phi)$, and is called the power of ϕ. Hence, a test ϕ of the risk probability α is said to be most powerful (MP) at $\theta' \in \Theta_1$ if $P_{\theta'}(\phi) \geq P_{\theta'}(\psi)$ holds for any test ψ of the risk probability α. Then, a test is said to be Uniformly Most Powerful (UMP) if it is MP at any $\theta' \in \Theta_1$.

B. p-values

In the hypothesis testing, we usually fixed our test before applying it to data. However, we sometimes focus on the minimum risk probability among tests in a class $\tilde{T}$ rejecting the hypothesis H_0 with a given data. This value is called the p-value, which depends on the observed data x as well as the subset Θ_0 to be rejected.

In fact, in order to define the p-value, we have to fix a class T of tests. Then, for x and Θ_0, p-value is defined as

$$\min_{\phi \in T : \phi(x)=1} \max_{\theta \in \Theta_0} P_\theta(\phi). \tag{7}$$

Since the p-value expresses the risk for rejecting the hypothesis H_0, Hence, this concept is useful for comparison among several designs of experiment.

Note that if we are allowed to choose any function ϕ as a test, the above minimum is attained by the function δ_x:

$$\delta_x(y) = \begin{cases} 0 & \text{if } y \neq x \\ 1 & \text{if } y = x. \end{cases} \tag{8}$$

In this case, the p-vale is $\max_{\theta \in \Theta_0} P_\theta(x)$. However, the function δ_x is unnatural as a test. Hence, we should fix a class of tests to define p-value.

IV. LIKELIHOOD TEST

A. Definition

In mathematical statistics, the likelihood ratio tests is often used as a class of standard tests[9]. This kind of tests often provide the UMP test in some typical cases. When both Θ_0 and Θ_1 consist of single elements as $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$, the likelihood ratio test $\phi_{\mathrm{LR},r}$ is defined as

$$\phi_{\mathrm{LR},r}(x) := \begin{cases} 0 & \text{if } P_{\theta_0}(x)/P_{\theta_1}(x) \geq r, \\ 1 & \text{if } P_{\theta_0}(x)/P_{\theta_1}(x) < r \end{cases}$$

where r is a constant, and the ratio $P_{\theta_0}(x)/P_{\theta_1}(x)$ is called the likelihood ratio. From the definition, any test ϕ satisfies

$$(rP_{\theta_1} - P_{\theta_0})(\phi_{\mathrm{LR},r}) \geq (rP_{\theta_1} - P_{\theta_0})(\phi). \qquad (9)$$

When a likelihood ratio test $\phi_{\mathrm{LR},r}$ satisfies

$$\alpha = P_{\theta_0}(\phi_{\mathrm{LR},r}), \qquad (10)$$

the test $\phi_{\mathrm{LR},r}$ is MP of level α. Indeed, when a test ϕ satisfies $P_{\theta_0}(\phi) \leq \alpha$,

$$\begin{aligned} &-\alpha + rP_{\theta_1}(\phi) = -P_{\theta_0}(\phi) + rP_{\theta_1}(\phi) \\ &\leq -P_{\theta_0}(\phi_{\mathrm{LR},r}) + rP_{\theta_1}(\phi_{\mathrm{LR},r}) = -\alpha + rP_{\theta_1}(\phi_{\mathrm{LR},r}). \end{aligned}$$

Hence, $1 - P_{\theta_1}(\phi) \geq 1 - P_{\theta_1}(\phi_{\mathrm{LR},r})$. This is known as Neyman-Pearson's fundamental lemma[15].

The likelihood ratio test is generalized to the cases where Θ_0 or Θ_1 has at least two elements as

$$\phi_{\mathrm{LR},r}(x) := \begin{cases} 0 & \text{if } \frac{\sup_{\theta \in \Theta_0} P_\theta(x)}{\sup_{\theta \in \Theta_1} P_\theta(x)} \geq r, \\ 1 & \text{if } \frac{\sup_{\theta \in \Theta_0} P_\theta(x)}{\sup_{\theta \in \Theta_1} P_\theta(x)} < r. \end{cases}$$

Usually, in order to guarantee a small risk probability, the likelihood ratio r is choosed as $r < 1$.

B. Monotone Likelihood Ratio Test

In cases where the hypothesis is one-sided, that is, the parameter space Θ is an interval of $\mathbb{R}$ and the hypothesis is given as

$$H_0 : \theta \geq \theta_0 \text{ versus } H_1 : \theta < \theta_0, \qquad (11)$$

we often use so-called interval tests for its optimality under some conditions as well as for its naturalness.

When the likelihood ratio $P_\theta(x)/P_\eta(x)$ is monotone increasing concerning x for any θ, η such that $\theta > \eta$, the likelihood ratio is called monotone. In this case, the likelihood ratio test $\phi_{\mathrm{LR},r}$ between P_{θ_0} and P_{θ_1} is UMP of level $\alpha := P_{\theta_0}(\phi_{\mathrm{LR},r})$, where θ_1 is an arbitrary element satisfying $\theta_1 < \theta_0$.

Indeed, many important examples satisfy this condition. Hence, it is convenient to give its proof here.

From the monotonicity, the likelihood ratio test $\phi_{\mathrm{LR},r}$ has the form

$$\phi_{\mathrm{LR},r}(x) = \begin{cases} 1 & x < x_0 \\ 0 & x \geq x_0 \end{cases} \qquad (12)$$

with a threshold value x_0. Since the monotonicity implies $P_{\theta_0}(\phi_{\mathrm{LR},r}) \geq P_\theta(\phi_{\mathrm{LR},r})$ for any $\theta \in \Theta_0$, it follows from Neyman Pearson Lemma that the likelihood ratio test $\phi_{\mathrm{LR},r}$ is MP of level α. From (12), the likelihood ratio test $\phi_{\mathrm{LR},r}$ is also a likelihood ratio test between P_{θ_0} and P_η, where η is another element satisfying $\eta < \theta_0$. Hence, the test $\phi_{\mathrm{LR},r}$ is also MP of level α.

From the above discussion, it is suitable to treat p-value based on the class of likelihood ratio tests. In this case, when we observe x_0, the p-value is equal to

$$\int_{-\infty}^{x_0} P_{\theta_0}(dx). \qquad (13)$$

C. One-Parameter Exponential Family

In mathematical statistics, exponential families are known as a class of typical statistical models[12]. A family of probability distributions $\{P_\theta | \theta \subset \Theta\}$ is called an exponential family when there exists a random variable x such that

$$P_\theta(x) := P_0(x) \exp(\theta x + g(\theta)), \qquad (14)$$

where $g(\theta) := -\log \int \exp(\theta x) P_0(dx)$.

It is known that this class of families includes, for example, the Poisson distributions, normal distributions, binomial distributions, etc. In this case, the likelihood ratio $\frac{\exp(\theta_0 x + g(\theta_0))}{\exp(\theta_1 x + g(\theta_1))} = \exp((\theta_0 - \theta_1)x + g(\theta_0) - g(\theta_1))$ is monotone concerning x for $\theta_0 > \theta_1$. Hence, the likelihood ratio test is UMP in the hypothesis (11). Note that this argument is valid even if we choose a different parameter if the family has a parameter satisfying (14).

For example, in the case of the normal distribution $P_\theta(x) = \frac{1}{\sqrt{2\pi V}} e^{-\frac{(x-\theta)^2}{2V}} = \frac{1}{\sqrt{2\pi V}} e^{-\frac{x^2}{2V} + \frac{\theta x}{V} - \frac{\theta^2}{2V}}$, the UMP test $\phi_{\mathrm{UMP},\alpha}$ of the level α is given as

$$\phi_{\mathrm{UMP},\alpha}(x) := \begin{cases} 1 & \text{if } x < \theta_0 - \epsilon_\alpha \sqrt{V} \\ 0 & \text{if } x \geq \theta_0 - \epsilon_\alpha \sqrt{V}, \end{cases} \qquad (15)$$

where

$$\Phi(-\epsilon_\alpha) = \alpha, \quad \Phi(\epsilon) = \int_{-\infty}^{\epsilon} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

The n-trial binomial distributions $P_p^n(k) = \binom{n}{k}(1-p)^{n-k}p^k$ are also an exponential family because another parameter $\theta := \log \frac{p}{1-p}$ satisfies that $P_p^n(k) = \binom{n}{k}\frac{1}{2^n} e^{\theta k + n \log \frac{e^\theta}{1+e^\theta}}$. Hence, in the case of the n-trial binomial distribution, the UMP test $\phi_{\mathrm{UMP},\alpha}^n$ of the level α is given as the randomized likelihood ratio test:

$$\phi_{\mathrm{UMP},\alpha}(k) := \begin{cases} 1 & \text{if } k < k_0 \\ \gamma & \text{if } k = k_0 \\ 0 & \text{if } k > k_0 \end{cases} \qquad (16)$$

where k_0 is the maximum value k' satisfying $\alpha \geq \sum_{k=0}^{k'-1} \binom{n}{k}(1-\theta)^{n-k}\theta^k$, and γ is defined as

$$\alpha = \gamma \binom{n}{k_0}(1-\theta)^{n-k_0}\theta^{k_0} + \sum_{k=0}^{k_0-1} \binom{n}{k}(1-\theta)^{n-k}\theta^k. \qquad (17)$$

6

Therefore, when k is observed, the p-value is $\sum_{k'=0}^{k} \binom{n}{k'}(1-\theta_0)^{n-k'}\theta_0^{k'}$.

When n is sufficiently large, the distribution $P_\theta^n(k)$ can be approximated by the normal distribution with variance $n(1-\theta)\theta$. Hence, the UMP test $\phi_{\mathrm{UMP},\alpha}^n$ of the level α is approximately given as

$$\phi_{\mathrm{UMP},\alpha}^n(k) = \begin{cases} 1 & \text{if } \frac{k}{n} < \theta_0 - \epsilon_\alpha\sqrt{\frac{\theta_0(1-\theta_0)}{n}} \\ 0 & \text{if } \frac{k}{n} \geq \theta_0 - \epsilon_\alpha\sqrt{\frac{\theta_0(1-\theta_0)}{n}}. \end{cases} \quad (18)$$

The p-value is also approximated to

$$\Phi\left(\frac{k - n\theta_0}{\sqrt{n(1-\theta_0)\theta_0}}\right). \quad (19)$$

The Poisson distributions $\mathrm{Poi}(\mu)$ are also an exponential family because another parameter $\theta := \log\mu$ satisfies $\mathrm{Poi}(\mu)(n) = \frac{1}{n}e^{\theta n - e^\theta}$. The UMP test $\phi_{\mathrm{UMP},\alpha}$ of the level α is characterized similarly to (16). When the threshold μ_0 is sufficiently large and the hypothesis is given

$$H_0 : \mu \geq \mu_0 \text{ versus } H_1 : \mu < \mu_0, \quad (20)$$

the UMP test $\phi_{\mathrm{UMP},\alpha}$ of the level α is approximately given as

$$\phi_{\mathrm{UMP},\alpha}(n) = \begin{cases} 1 & \text{if } n < \mu_0 - \epsilon_\alpha\sqrt{\mu_0} \\ 0 & \text{if } n \geq \mu_0 - \epsilon_\alpha\sqrt{\mu_0}. \end{cases} \quad (21)$$

The p-value is also approximated to

$$\Phi\left(\frac{n - \mu_0}{\sqrt{\mu_0}}\right), \quad (22)$$

Next, we consider testing the following hypothesis in the case of the binomial Poisson distribution $\mathrm{Poi}(\mu_1, \mu_2)$:

$$H_0 : \frac{\mu_1}{\mu_1 + \mu_2} \geq \theta_0 \text{ versus } H_1 : \frac{\mu_1}{\mu_1 + \mu_2} < \theta_0. \quad (23)$$

In this case, as is shown at (32) and (31) in Section IV D, the likelihood ratio test $\phi_{\mathrm{LR},r}$ is characetrized by the likelihood ratio test of the binomial distributions as

$$\phi_{\mathrm{LR},r}(n_1, n_2) = \phi_{\mathrm{LR},r}^{n_1+n_2}(n_1). \quad (24)$$

Hence, it is suitable to employ the likelihood ratio test $\phi_{\mathrm{LR},l=\alpha}(n_1, n_2) = \phi_{\mathrm{UMP},\alpha}^{n_1+n_2}(n_1)$ with the level α. This is because the conditional distribution $\frac{\mathrm{Poi}(\mu_1,\mu_2)(n_1,n_2)}{\sum_{k'=0}^{n_1+n_2}\mathrm{Poi}(\mu_1,\mu_2)(k',n_1+n_2-k')}$ is equal to the binomial distribution $P_{\frac{\mu_1}{\mu_1+\mu_2}}^{n_1+n_2}(n_1)$. Therefore, when we observe n_1, n_2, the p-value of this class of likelihood ratio tests is equal to $\sum_{k=0}^{n_1} \binom{n_1+n_2}{k}\theta_0^k(1-\theta_0)^{n_1+n_2-k}$.

When the total number $n_1 + n_2$ is sufficiently large, the test $\phi_{\mathrm{LR},l=\alpha}$ of the level α is approximately given as

$$\phi_{\mathrm{LR},l=\alpha}(n_1, n_2) := \begin{cases} 1 & \text{if } \frac{n_1}{n_1+n_2} < \theta_0 - \epsilon_\alpha\sqrt{\frac{\theta_0(1-\theta_0)}{n_1+n_2}} \\ 0 & \text{if } \frac{n_1}{n_1+n_2} \geq \theta_0 - \epsilon_\alpha\sqrt{\frac{\theta_0(1-\theta_0)}{n_1+n_2}}. \end{cases}$$
$$(25)$$

The p-value is also approximated to

$$\Phi\left(\frac{n_1 - (n_1 + n_2)\theta_0}{\sqrt{(n_1 + n_2)(1-\theta_0)\theta_0}}\right). \quad (26)$$

D. Multi-parameter case

In the one-parameter case, UMP tests can be often characterized by likelihood ratiotests. However, in the multi-parameter case, this type characterization is impossible generally, and the UMP test does not always exist. In this case, we have to choose our test among non-UMP tests. One idea is choosing our test among likelihood ratio tests because likelihood ratio tests always exist and we can expect that these tests have good performances. Generally, it is not easy to give an explicit form of the likelihood ratio test. When the family is a multiparameter exponential family, the likelihood ratio test has a simple form. A family of probability distributions $\{P_{\vec{\theta}}|\vec{\theta} = (\theta^1, \ldots, \theta^m) \in \mathbb{R}^m\}$ is called an m-parameter exponential family when there exists m-dimensional random variable $\vec{x} = (x_1, \ldots, x_m)$ such that

$$P_{\vec{\theta}}(\vec{x}) := P_0(\vec{x})\exp(\vec{\theta} \cdot \vec{x} + g(\vec{\theta})),$$

where $g(\vec{\theta}) := -\log\int\exp(\vec{\theta} \cdot \vec{x})P_0(d\vec{x})$. However, this form is not sufficiently simple because its rejection region is given by the a nonlinear constraint. Hence, a test with a simpler form is required. In the following, we discuss the likelihood ratio test in the case of multi-nomial Poisson distribution. After this discussion, we propose an alternative test.

In an m-parameter exponential family, the likelihood ratio test $\phi_{\mathrm{LR},r}$ has the form

$$\phi_{\mathrm{LR},r}(\vec{x}) = \begin{cases} 0 \text{ if } & \inf_{\vec{\theta}_1 \in \Theta_1} D(P_{\vec{\theta}(\vec{x})}\|P_{\vec{\theta}_1}) \\ & -\inf_{\vec{\theta}_0 \in \Theta_0} D(P_{\vec{\theta}(\vec{x})}\|P_{\vec{\theta}_0}) \geq \log r \\ 1 \text{ if } & \inf_{\vec{\theta}_1 \in \Theta_1} D(P_{\vec{\theta}(\vec{x})}\|P_{\vec{\theta}_1}) \\ & -\inf_{\vec{\theta}_0 \in \Theta_0} D(P_{\vec{\theta}(\vec{x})}\|P_{\vec{\theta}_0}) < \log r, \end{cases}$$
$$(27)$$

where the divergence $D(P_{\vec{\eta}}\|P_{\vec{\theta}})$ is defined as

$$D(P_{\vec{\eta}}\|P_{\vec{\theta}}) := \int \log\frac{P_{\vec{\eta}}(\vec{x}')}{P_{\vec{\theta}}(\vec{x}')}P_{\vec{\eta}}(d\vec{x}')$$
$$= (\vec{\eta} - \vec{\theta})\int \vec{x}P_{\vec{\eta}}(d\vec{x}) + g(\vec{\eta}) - g(\vec{\theta}),$$

and $\vec{\theta}(\vec{x})$ is defined by [12]

$$\int \vec{x}'P_{\vec{\theta}(\vec{x})}(d\vec{x}') = \vec{x}. \quad (28)$$

This is because the logarithm of the likelihood function

is calculated as

$$\log \frac{\sup_{\vec{\theta}_0 \in \Theta_0} P_{\vec{\theta}_0}(\vec{x})}{\sup_{\vec{\theta}_1 \in \Theta_1} P_{\vec{\theta}_1}(\vec{x})}$$

$$= \sup_{\vec{\theta}_0 \in \Theta_0} \inf_{\vec{\theta}_1 \in \Theta_1} \log \frac{P_{\vec{\theta}_0}(\vec{x})}{P_{\vec{\theta}_1}(\vec{x})}$$

$$= \sup_{\vec{\theta}_0 \in \Theta_0} \inf_{\vec{\theta}_1 \in \Theta_1} (\vec{\theta}_0 - \vec{\theta}_1) \cdot \vec{x} + g(\vec{\theta}_0) - g(\vec{\theta}_0)$$

$$= \sup_{\vec{\theta}_0 \in \Theta_0} \inf_{\vec{\theta}_1 \in \Theta_1} (\vec{\theta}_0 - \vec{\theta}_1) \cdot \int \vec{x}' P_{\vec{\theta}(\vec{x})}(d\vec{x}') + g(\vec{\theta}_0) - g(\vec{\theta}_0)$$

$$= \sup_{\vec{\theta}_0 \in \Theta_0} \inf_{\vec{\theta}_1 \in \Theta_1} D(P_{\vec{\theta}(\vec{x})} \| P_{\vec{\theta}_1}) - D(P_{\vec{\theta}(\vec{x})} \| P_{\vec{\theta}_0})$$

$$= \inf_{\vec{\theta}_1 \in \Theta_1} D(P_{\vec{\theta}(\vec{x})} \| P_{\vec{\theta}_1}) - \inf_{\vec{\theta}_0 \in \Theta_0} D(P_{\vec{\theta}(\vec{x})} \| P_{\vec{\theta}_0}).$$

In addition, $\vec{\theta}(\vec{x})$ coincides with the MLE when $\vec{x}$ is observed. Hence, when $\Theta = \Theta_0 \cup \Theta_1$, the likelihood ratio test with the ratio $r < 1$ is given by the rejection region:

$$\left\{ \vec{x} \, \middle| \, \vec{\theta}(\vec{x}) \in \Theta_1, \inf_{\vec{\theta}_0 \in \Theta_0} D(P_{\vec{\theta}(\vec{x})} \| P_{\vec{\theta}_0}) \geq -\log r \right\}. \quad (29)$$

In the case of the multi-nomial Poisson distributions $\text{Poi}(\vec{\mu})(\vec{k}) := e^{-\sum_{i=1}^{l} \mu_i} \frac{\mu_1^{k_1} \cdots \mu_l^{k_m}}{k_1! \cdots k_m!}$, which is an exponential family, the divergence is calculated as

$$D(\text{Poi}(\vec{\mu}) \| \text{Poi}(\vec{\mu}'))$$

$$= \sum_{i=1}^{m} (\mu_i' - \mu_i) + \sum_{i=1}^{m} \mu_i \log \frac{\mu_i}{\mu_i'} \quad (30)$$

$$= (\sum_{i=1}^{m} \mu_i') - (\sum_{i=1}^{m} \mu_i)(\sum_{i=1}^{m} \mu_i) \log \frac{\sum_{i=1}^{m} \mu_i}{\sum_{i=1}^{m} \mu_i'}$$

$$+ (\sum_{i=1}^{m} \mu_i) D(\frac{\vec{\mu}}{\sum_{i=1}^{m} \mu_i} \| \frac{\vec{\mu}'}{\sum_{i=1}^{m} \mu_i'}), \quad (31)$$

where $D(\vec{p} \| \vec{p}')$ is the divergence between the multinomial distributions $\vec{p}$ and $\vec{p}'$.

When the hypothesis is given by (23) and $\frac{n_1}{n_1+n_2} \leq \theta_0$, we have

$$\log \frac{\sup_{\vec{\theta}_0 \in \Theta_0} P_{\vec{\theta}_0}(n_1, n_2)}{\sup_{\vec{\theta}_1 \in \Theta_1} P_{\vec{\theta}_1}(n_1, n_2)}$$

$$= (n_1 + n_2) D(P_{\frac{n_1}{n_1+n_2}} \| P_{\theta_0}) = D(P_{\frac{n_1}{n_1+n_2}}^{n_1+n_2} \| P_{\theta_0}^{n_1+n_2}), \quad (32)$$

where P_θ is the binomial distribution with one observation and P_θ^n is the binomial distribution with n observations. Then, the likelihood ratio test is given by the likelihood ratio test of the binomial distributions.

In the following, we treat two hypotheses given as

$$H_0 : \vec{w} \cdot \vec{\theta} \geq c_0 \text{ versus } H_1 : \vec{w} \cdot \vec{\theta} < c_0, \quad (33)$$

with the condition $w_i \geq 0$, Using the formula (30), and (33), we can calculate the likelihood ratio test for a given

ratio r. Now, we calculate the p-value concerning the class of likelihood ratio tests when we observe the data $k_1, \ldots, k_m$. When $\vec{w} \cdot \vec{k} < c_0$, this p-value is equal to

$$\max_{\vec{w} \cdot \vec{\mu}' = c_0} \text{Poi}(\vec{\mu}')(\mathcal{A}_{R(\vec{k})}), \quad (34)$$

where

$$\mathcal{A}_R := \left\{ \vec{k}' \, \middle| \, \begin{array}{l} \min_{\vec{w} \cdot \vec{\mu} = c_0} \sum_{i=1}^{m} (\mu_i - k_i') + \sum_{i=1}^{m} k_i' \log \frac{k_i'}{\mu_i} \geq R \\ \vec{w} \cdot \vec{k}' < c_0 \end{array} \right\},$$

$$R(\vec{k}) := \min_{\vec{w} \cdot \vec{\mu} = c_0} \sum_{i=1}^{m} (\mu_i - k_i) + \sum_{i=1}^{m} k_i \log \frac{k_i}{\mu_i}$$

because the minimum R satisfying $\vec{k} \in \mathcal{A}_R$ is $R(\vec{K})$. Since the calculation of (34) is not so easy, we consider its upper bound. For this purpose, we define the set $\mathcal{B}_R$ as

$$\mathcal{B}_R := \left\{ \vec{k}' \, \middle| \, \sum_{i=1}^{m} \frac{k_i'}{\tilde{\mu}_i(R)} \leq 1 \right\}, \quad (35)$$

where $\tilde{\mu}_i(R)$ are defined as follows:

$$\frac{c_0}{w_i} - \tilde{\mu}_i(R) + \tilde{\mu}_i(R) \log \frac{\tilde{\mu}_i(R) w_i}{c_0} = R \text{ if } R \leq R_{0,i} \quad (36)$$

$$\frac{c_0}{w_M} + \tilde{\mu}_i(R) \log \frac{w_M - w_i}{w_M} = R \text{ if } R > R_{0,i}, \quad (37)$$

where $w_M := \max_i w_i$ and $R_{0,i} := \frac{c_0}{w_M} + \frac{c_0(w_M - w_i)}{w_i w_M} \log \frac{w_M - w_i}{w_M}$. Note that $\tilde{\mu}_i(R)$ is a monotone decreasing function of R. As is shown in Appendix D,

$$\mathcal{A}_R \subset \mathcal{B}_R. \quad (38)$$

Then, the p-value concerning likelihood ratio tests is upperly bounded by

$$\max_{\vec{w} \cdot \vec{\mu}' = c_0} \text{Poi}(\vec{\mu}')(\mathcal{B}_{R(\vec{k})}). \quad (39)$$

However, it is difficult to choose the likelihood r such that the p-value is equal to a given risk probability α because the set $\mathcal{A}_R$ is defined by a non-linear constraint. In order to resolve this problem, we propose to modify the likelihood ratio test by using the set $\mathcal{B}_R$ instead of the set $\mathcal{A}_R$ because $\mathcal{B}_R$ is defined by a linear constraint while $\mathcal{A}_R$ is by a non-linear constraint. That is, we define the modified test $\phi_{\text{mod},R}$ as the test with the rejection region $\mathcal{B}_R$. Among this kind of tests, we can choose the test $\phi_{\text{mod},R_\alpha}$ with the risk probability α by choosing R_α in the following way:

$$\max_{\vec{w} \cdot \vec{\mu}' = c_0} \text{Poi}(\vec{\mu}')(\mathcal{B}_{R_\alpha}) = \alpha. \quad (40)$$

Indeed, the calculation of the probability $\text{Poi}(\vec{\mu}')(\mathcal{B}_R)$ is easier than that of the probability $\text{Poi}(\vec{\mu}')(\mathcal{A}_R)$ because of the linearity of the constraint condition of $\mathcal{B}_R$.

Next, we calculate the p-value of the set of the modified tests $\{\phi_{\mathrm{mod},\alpha}\}_\alpha$. For an observed data $\vec{k}$, we choose $R'(\vec{k})$ as R' satisfying

$$\sum_{i=1}^{m} \frac{k_i}{\tilde{\mu}_i(R')} = 1. \tag{41}$$

The LHS is monotone increasing for R' because each $\tilde{\mu}_i(R')$ is monotone decreasing for R'. Thus, $R'(\vec{k})$ is the maximum R' such that $\vec{k} \in \mathcal{B}_{R'}$. Then, the p-value is equal to $\max_{\vec{w}\cdot\vec{\mu}'=c_0} \mathrm{Poi}(\vec{\mu}')(\mathcal{B}_{R'(\vec{k})})$. Further, the relation (38) implies $\vec{k} \in \mathcal{B}_{R(\vec{k})}$. Hence, $R(\vec{k}) \leq R'(\vec{k})$, which implies $\mathcal{B}_{R(\vec{k})} \supset \mathcal{B}_{R'(\vec{k})}$. Therefore, the p-value $\max_{\vec{w}\cdot\vec{\mu}'=c_0} \mathrm{Poi}(\vec{\mu}')(\mathcal{B}_{R'(\vec{k})})$ concerning the modified tests $\{\phi_{\mathrm{mod},\alpha}\}_\alpha$ is smaller than the upper bound $\max_{\vec{w}\cdot\vec{\mu}'=c_0} \mathrm{Poi}(\vec{\mu}')(\mathcal{B}_{R(\vec{k})})$ of p-value concerning the likelihood ratio tests. This test ϕ_{mod} coincides with the likelihood ratio test in the one-parameter case.

V. ASYMPTOTIC THEORY

A. Fisher information

Assume that the data $x_1, \ldots, x_n$ obeys the identical and independent distribution of the same distribution family p_θ and n is sufficiently large. When the true parameter θ is close to θ_0, it is known that the meaningful information for θ is essentially given as the random variable $\frac{1}{n}\sum_{i=1}^{n} l_{\theta_0}(x_i)$, where the logarithmic derivative $l_{\theta_0}(x_i)$ is defined by

$$l_\theta(x) := \frac{d \log p_\theta(x)}{d\theta}. \tag{42}$$

In this case, the random variable $\frac{1}{n}\sum_{i=1}^{n} l_{\theta_0}(x_i)$ can be approximated by the normal distribution with the expectation value $\theta - \theta_0$ and the variance $\frac{1}{nJ_{\theta_0}}$, where the Fisher information J_θ is defined as $J_\theta := \int (l_\theta(x))^2 P_\theta(dx)$. Hence, the testing problem can be approximated by the testing of this normal distribution family [9, 12]. That is, the quality of testing is approximately evaluated by the Fisher information J_{θ_0} at the threshold θ_0.

In the case of Poisson distribution family $\mathrm{Poi}(\theta t)$, the parameter θ can be estimated by $\frac{X}{t}$. The asymptotic case corresponds to the case with large t. In this case, Fisher information is $\frac{t}{\theta}$. When X obeys the unknown Poisson distribution family $\mathrm{Poi}(\theta t)$, the estimation error $\frac{X}{t} - \theta$ is close to the normal distribution with the variance $\frac{\theta}{t}$, i.e., $\sqrt{t}(\frac{X}{t} - \theta)$ approaches to the random variables obeying the normal distribution with variance θ. That is, Fisher information corresponds to the inverse of variance of the estimator.

This approximation can be extended to the multi-parameter case $\{p_\theta | \theta \in \mathbb{R}^m\}$. Similarly, it is known that the testing problem can be approximated by the testing

of the normal distribution family with the covariance matrix $(nJ_\theta)^{-1}$, where the Fisher information matrix $J_{\theta;i,j}$ is given by

$$J_{\theta;i,j} := \int l_{\theta;i}(x) l_{\theta;j}(x) P_\theta(dx), \tag{43}$$

$$l_{\theta;i}(x) := \frac{\partial \log p_\theta(x)}{\partial \theta^i}. \tag{44}$$

When the hypotheses is given by (11), the testing problem can be approximated by the testing of the normal distribution family with variance $\frac{\vec{w}\cdot J_{\theta_0}^{-1}\vec{w}}{n}$,

Indeed, the same fact holds for the multinomial Poisson distribution family $\mathrm{Poi}(t\vec{\mu})$. When the random variable X_j is the i-th random variable, the random variable $\sum_{j=1}^{m} \frac{\lambda_j}{\sqrt{t}}(X_j - \mu_j)$ converges to the random variable obeying the normal distribution with the variance $\sum_{j=1}^{m} \lambda_j^2 \mu_j$ in distribution:

$$\sum_{j=1}^{m} \frac{\lambda_j}{\sqrt{t}}(X_j - \mu_j) \xrightarrow{d} \sum_{j=1}^{m} \lambda_j^2 \mu_j. \tag{45}$$

This convergence is compact uniform concerning the parameter $\vec{\mu}$. In this case, the Fisher information matrix J_μ is the diagonal matrix with the diagonal elements $(\frac{t}{\mu_1}, \ldots, \frac{t}{\mu_m})$. When our distribution family is given as a subfamily $\mathrm{Poi}(t\mu_1(\theta), \ldots, t\mu_m(\theta))$, the Fisher information matrix is $\mathcal{A}_\theta^t J_{\mu(\theta)} \mathcal{A}_\theta$, where $\mathcal{A}_{\theta;i,j} = \frac{\partial \mu_j}{\partial \theta_i}$. Hence, when the hypotheses is given by (33), the testing problem can be approximated by the testing of the normal distribution family with variance

$$\vec{w} \cdot (\mathcal{A}_\theta^t J_{\mu(\theta)} \mathcal{A}_\theta)^{-1}\vec{w}. \tag{46}$$

In the following, we call this value Fisher information. Based on this value, the quality can be compared when we have several testing schemes.

B. Multi-parametric Poisson distribution

In the following, we treat testing of the hypothesis (33) in the multinomial Poisson distribution $\mathrm{Poi}(\vec{\mu})$ by using normal approximation. In this case, by using $\tilde{\mu}_i$ defined in (36) and (37), the upper bound (39) of the p-value concerning the likelihood ratio tests is approximated to

$$\max_{w\cdot\mu'=c_0} \Phi\left(\frac{1 - \sum_{j=1}^{m} \frac{\mu_j'}{\tilde{\mu}_j(R(\vec{k}))}}{\sqrt{\sum_{i=1}^{m} \frac{\mu_i'}{\tilde{\mu}_i(R(\vec{k}))^2}}}\right)$$

$$= \Phi\left(\max_{w\cdot\mu'=c_0} \frac{1 - \sum_{j=1}^{m} \frac{\mu_j'}{\tilde{\mu}_j(R(\vec{k}))}}{\sqrt{\sum_{i=1}^{m} \frac{\mu_i'}{\tilde{\mu}_i(R(\vec{k}))^2}}}\right),$$

because this convergence (45) is compact uniform concerning the parameter $\vec{\mu}$. Letting $x_i(R) = \frac{c_0}{w_i \tilde{\mu}_i(R)} - 1$ and $y_i(R) = \frac{c_0}{w_i \tilde{\mu}_i(R)^2}$, we have

$$\max_{w \cdot \mu' = c_0} \frac{1 - \sum_{j=1}^m \frac{\mu'_j}{\tilde{\mu}_j(R(\vec{k}))}}{\sqrt{\sum_{i=1}^m \frac{\mu'_i}{\tilde{\mu}_i(R(\vec{k}))^2}}} = \max_{(x,y) \in Co(R(\vec{k}))} \frac{-x}{\sqrt{y}}, \quad (47)$$

where $Co(R)$ is the convex hull of $(x_1(R), y_1(R)), \ldots, (x_m(R), y_m(R))$. As is shown in Appendix E, this value is simplified to

$$-\min_{i,j} z_{i,j}(R(\vec{k})), \quad (48)$$

where

$$z_{i,j}(R) := \begin{cases} \frac{x_i(R)}{\sqrt{y_i(R)}} & \text{if } \frac{2x_j(R)y_i(R)}{x_i(R)y_i(R) + x_i(R)y_j(R)} \geq 1 \\ \frac{x_j(R)}{\sqrt{y_j(R)}} & \text{if } \frac{2x_i(R)y_j(R)}{x_j(R)y_j(R) + x_j(R)y_i(R)} \geq 1 \\ \tilde{z}_{i,j}(R) & \text{otherwise,} \end{cases} \quad (49)$$

where

$$\tilde{z}_{i,j}(R) := \frac{2(x_i(R)x_j(R)(y_i(R) + y_j(R)) - x_i(R)^2 y_j(R) - x_j(R)^2 y_i(R))}{\sqrt{(x_i(R) - x_j(R))(y_i(R) - y_j(R))}\sqrt{x_i(R)y_j(R)^2 + x_j(R)y_i(R)^2 - y_i(R)y_j(R)(x_i(R) + x_j(R))}} \quad (50)$$

That is, our upper bound of p-value concerning the likelihood ratio tests is given by

$$\Phi(-\min_{i,j} z_{i,j}(R(\vec{k}))). \quad (51)$$

Next, we approximately calculate the test with the risk probability α proposed in sectionIV D. First, we choose R_α by

$$-\min_{i,j} z_{i,j}(R_\alpha) = \Phi^{-1}(\alpha). \quad (52)$$

Then, our test is given by the rejection region $\mathcal{B}_{R_\alpha}$. Using the same discussion, the p-value concerning the proposed tests is equal to

$$\Phi(-\min_{i,j} z_{i,j}(R'(\vec{k}))). \quad (53)$$

VI. MODIFICATION OF VISIBILITY

In the following sections, we apply the discussions in sections III - V to the hypothesis (3). That is, we consider how to reject the null hypothesis $H_0 : F \leq F_0$ with a certain risk probability α.

In the usual visibility, we usually measure the coincidence events only in the one direction or two directions. However, in this method, the number of the counts of coincidence events be reflected not only by the fidelity but also by the direction of difference between the true state of target maximally entangled state. In order to remove the bias based on such a direction, we propose to measure the counts of the coincidence vectors $|HH\rangle, |VV\rangle, |DD\rangle, |XX\rangle, |RL\rangle$, and $|LR\rangle$, which corresponds to the coincidence events, and the counts of the

anti-coincidence vectors $|HV\rangle, |VH\rangle, |DX\rangle, |XD\rangle, |RR\rangle$, and $|LL\rangle$, which corresponds to the anti-coincidence events. The former corresponds to the the minimum values in the usual visibility, and the later does to the minimum values in the usual visibility. In this paper, we call this proposed method the modified visibility method. Using this method, we can test the fidelity between the maximally entangled state $|\Phi^{(+)}\rangle\langle\Phi^{(+)}|$ and the given state σ, using the total number of counts of the coincidence events (the total count on coincidence event) n_1 and the total number of counts of the anti-coincidence events (the total count on anti-coincidence events) n_2 obtained by measuring on all the vectors with the time $\frac{t}{12}$. When the dark count is negligible, the total count on coincidence events n_1 obeys $\text{Poi}(\lambda \frac{2F+1}{12} t)$, and the count on total anti-coincidence events n_2 obeys the distribution $\text{Poi}(\lambda \frac{2-2F}{12} t)$. These expectation values μ_1 and μ_2 are given as $\mu_1 = \lambda \frac{2F+1}{12} t$ and $\mu_2 = \lambda \frac{2-2F}{12} t$. Hence, Fisher information matrix concerning the parameters F and λ is

$$\begin{pmatrix} \lambda(\frac{t}{3(2F+1)} + \frac{t}{3(2-2F)}) & 0 \\ 0 & \frac{\frac{2F+1}{12}t + \frac{2-2F}{12}t}{\lambda} \end{pmatrix}, \quad (54)$$

where the first element corresponds to the parameter F and the second one does to the parameter λ. Then, we can apply the test ϕ_{LR} given in the end of subsection IV C. That is, based on the ratio $\frac{\mu_2}{\mu_1 + \mu_2} = \frac{2}{3}(1 - F)$, we estimate the fidelity using the ratio $\frac{n_2}{n_1 + n_2}$ as $\hat{F}(n_1, n_2) = 1 - \frac{3}{2} \frac{n_2}{n_1 + n_2}$. Based on the discussion in subsection V A, its variance is asymptotically equal to

$$\frac{1}{\lambda(\frac{t}{3(2F+1)} + \frac{t}{3(2-2F)})} = \frac{(2F+1)(2-2F)}{\lambda t}. \quad (55)$$

Hence, similarly to the visibility, we can check the fidelity by using this ratio.

Indeed, when we consider the distribution under the condition that the total count $n_1 + n_2$ is fixed to n, the random variable n_2 obeys the binomial distribution with the average value $\frac{2}{3}(1-F)n$. Hence, we can apply the likelihood ratio test of the binomial distribution. In this case, by the approximation to the normal distribution, the likelihood ratio test with the risk probability α is almost equal to the test with the rejection region: $\{(n_1, n_2) | \frac{n_2}{n_1+n_2} \le \frac{2}{3}(1-F_0) + \Phi^{-1}(\alpha)\sqrt{\frac{(2-2F_0)(1+2F_0)}{9(n_1+n_2)}}\}$ concerning the null hypothesis $H_0 : F \le F_0$. The p-value of this kind of tests is $\Phi(\frac{n_2(2F_0+1) - n_1(2-2F_0)}{\sqrt{(n_1+n_2)(2F_0+1)(2-2F_0)}})$.

VII.   DESIGN I (λ: UNKNOWN, ONE STAGE)

In this section, we consider the problem of testing the fidelity between the maximally entangled state $|\Phi^{(+)}\rangle\langle\Phi^{(+)}|$ and the given state σ by performing three kinds of measurement, coincidence, anti-coincidence, and total flux, with the times t_1, t_2 and t_3, respectively. When the dark count is negligible, the data (n_1, n_2, n_3) obeys the multinomial Poisson distribution $\text{Poi}(\lambda\frac{2F+1}{6}t_1, \lambda\frac{2-2F}{6}t_2, \lambda t_3)$ with the assumption that the parameter λ is unknown. In this problem, it is natural to assume that we can select the time allocation with the constraint for the total time $t_1 + t_2 + t_3 = t$.

The performance of the time allocation (t_1, t_2, t_3) can evaluated by the variance (46). The Fisher information matrix concerning the parameters F and λ is

$$\begin{pmatrix} \lambda(\frac{2t_1}{3(2F+1)} + \frac{2t_2}{3(2-2F)}) & \frac{t_1-t_2}{3} \\ \frac{t_1-t_2}{3} & \frac{\frac{2F+1}{6}t_1 + \frac{2-2F}{6}t_2 + t_3}{\lambda} \end{pmatrix}, \quad (56)$$

where the first element corresponds to the parameter F and the second one does to the parameter λ. Then, the asymptotic variance (46) is calculated as

$$\frac{\frac{2F+1}{6}t_1 + \frac{2-F}{6}t_2 + t_3}{\lambda\left((\frac{2F+1}{6}t_1 + \frac{2-F}{6}t_2 + t_3)(\frac{2t_1}{3(2F+1)} + \frac{2t_2}{3(2-2F)}) - (\frac{t_1-t_2}{3})^2\right)}. \quad (57)$$

We optimize the time allocation by minimizing the variance (57). We perform the minimization by maximizing the inverse: $\lambda\left(\frac{2t_1}{3(2F+1)} + \frac{2t_2}{3(2-2F)} - \frac{(\frac{t_1-t_2}{3})^2}{\frac{2F+1}{6}t_1 + \frac{2-2F}{6}t_2 + t_3}\right)$. Applying Lemmas 1 and 2 shown in Appendix A to the case of $a = \frac{2}{3(2F+1)}$, $b = \frac{2}{3(2-2F)}$, $c = \frac{2F+1}{6}$, $d = \frac{2-2F}{6}$, we obtain

(i)

$$\lambda \max_{t_1+t_3=t} \frac{2t_1}{3(2F+1)} - \frac{(\frac{t_1}{3})^2}{\frac{2F+1}{6}t_1 + t_3} = \frac{2\lambda t}{3(2F+1)(1 + \sqrt{\frac{2F+1}{6}})^2} \quad (58)$$

(ii)

$$\lambda \max_{t_2+t_3=t} \frac{2t_2}{3(2-2F)} - \frac{(\frac{t_2}{3})^2}{\frac{2-2F}{6}t_2 + t_3} = \frac{2\lambda t}{3(2-2F)(1 + \sqrt{\frac{2-2F}{6}})^2} \quad (59)$$

and

(iii) $\lambda \max_{t_1+t_2=t} \frac{2t_1}{3(2F+1)} + \frac{2t_2}{3(2-2F)} - \frac{(\frac{t_1-t_2}{3})^2}{\frac{2F+1}{6}t_1 + \frac{2-2F}{6}t_2}$

$$= \frac{\lambda(\frac{1}{3}\sqrt{\frac{2-2F}{2F+1}} + \frac{1}{3}\sqrt{\frac{2F+1}{2-2F}})^2 t}{(\sqrt{\frac{2F+1}{6}} + \sqrt{\frac{2-2F}{6}})^2}$$

$$= \frac{6\lambda t}{(2F+1)(2-2F)(\sqrt{2F+1} + \sqrt{2-2F})^2}. \quad (60)$$

Then, these relations give the optimal time allocations between (i) coincidence and total flux measurements, (ii) anti-coincidence and total flux measurements, and (iii) coincidence and anti-coincidence measurements, respectively. The ratio of (60) to (58) is equal to

$$\frac{3(\sqrt{6} + \sqrt{2F+1})^2}{2(2-2F)(\sqrt{2F+1} + \sqrt{2-2F})^2} > 1, \quad (61)$$

as shown in Appendix B. That is, the optimal measurement using the coincidence and the anti-coincidence always provides better test than that using the coincidence and the total flux. Hence, we compare (ii) with (iii), and obtain

$$\max_{t_1+t_2+t_3=t} \lambda\Big(\frac{2t_1}{3(2F+1)} + \frac{2t_2}{3(2-2F)}$$

$$- \frac{(\frac{t_1-t_2}{3})^2}{\frac{2F+1}{6}t_1 + \frac{2-2F}{6}t_2 + t_3}\Big)$$

$$= \begin{cases} \frac{4\lambda t}{(2-2F)(\sqrt{6}+\sqrt{2-2F})^2} & \text{if } F_1 < F \le 1 \\ \frac{6\lambda t}{(2F+1)(2-2F)(\sqrt{2F+1}+\sqrt{2-2F})^2} & \text{if } 0 \le F \le F_1, \end{cases} \quad (62)$$

where the critical point $F_1 < 1$ is defined by

$$\frac{2(2F_1+1)(\sqrt{2F_1+1} + \sqrt{2-2F_1})^2}{3(\sqrt{6} + \sqrt{2-2F_1})^2} = 1. \quad (63)$$

The approximated value of the critical point F_1 is 0.899519. The equation (62) is derived in Appendix C.

Fig. 2 shows the ratio of the optimal Fisher information based on the anti-coincidence and total flux measurements to that based on the coincidence and anti-coincidence measurements. When $F_1 \le F \le 1$, the maximum Fisher information is attained by $t_1 = 0$, $t_2 = \frac{\sqrt{6}}{(\sqrt{6}+\sqrt{2(1-F)})}t$, $t_3 = \frac{\sqrt{2(1-F)}}{\sqrt{6}+\sqrt{2(1-F)}}t$. Otherwise, the maximum is attained by $t_1 = \frac{\sqrt{2-2F}}{\sqrt{2F+1}+\sqrt{2-2F}}t$, $t_2 = \frac{\sqrt{2F+1}}{\sqrt{2F+1}+\sqrt{2-2F}}t$, $t_3 = 0$. The optimal time allocation
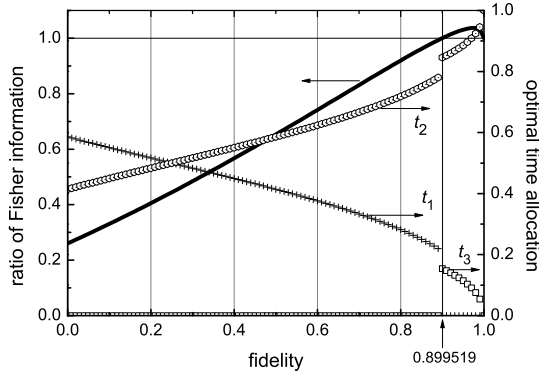
FIG. 2: The ratio of the optimal Fisher information (solid line) and the optimal time allocation as a function of the fidelity F. The measurement time is divided into three periods: coincidence t_1 (plus signs), anti-coincidence t_2 (circles), and total flux t_3 (squares), which are normalized as $t_1+t_2+t_3 = 1$ in the plot.

shown in Fig. 2 implies that we should measure the counts on the anti-coincidence vectors preferentially over other vectors.

The optimal asymptotic variance is $\frac{(2F+1)(2-2F)(\sqrt{2-2F}+\sqrt{1+2F})^2}{6\lambda t}$ when the threshold F_0 is less than the critical point F_1. This asymptotic variance is much better than that obtained by the modified visibility method. The ratio of the optimal asymptotic variance is given by

$$\frac{(\sqrt{2-2F}+\sqrt{1+2F})^2}{6} < 1. \qquad (64)$$

In the following, we give the optimal test of level α in

the hypothesis testing (6). Assume that the threshold F_0 is less than the critical point F_1. In this case, we can apply testing of the hypothesis (23). First, we measure the count on the coincidence vectors for a period of $t_1 = \frac{t\sqrt{2-2F_0}}{\sqrt{2F_0+1}+\sqrt{2-2F_0}}$, to obtain the total count n_1. Then, we measure the count on the anti-coincidence vectors for a period of $t_2 = \frac{t\sqrt{2F_0+1}}{\sqrt{2F_0+1}+\sqrt{2-2F_0}}$ to obtain the total count n_2. Note that the optimal time allocation depends on the threshold of our hypothesis. Finally, we apply the UMP test of α of the hypothesis:

$$H_0 : p \geq \frac{\sqrt{2-2F_0}}{\sqrt{2-2F_0}+\sqrt{1+2F_0}} \quad \text{versus} \quad H_1 : p < \frac{\sqrt{2-2F_0}}{\sqrt{2-2F_0}+\sqrt{1+2F_0}}$$

with the binomial distribution family $P_p^{n_1+n_2}$ to the data n_1. In this case, the likelihood ratio test with the risk probability α is almost equal to the test with the rejection region: $\{(n_1,n_2)|\frac{n_2}{n_2+n_1} \leq \frac{\sqrt{2-2F_0}}{\sqrt{2F_0+1}+\sqrt{2-2F_0}} + \frac{\Phi^{-1}(\alpha)}{\sqrt{2F_0+1}+\sqrt{2-2F_0}}\sqrt{\frac{\sqrt{2-2F_0}\sqrt{2F_0+1}}{n_1+n_2}}\}$ concerning the null hypothesis $H_0 : F \leq F_0$. The p-value of this kind of tests is $\Phi(\frac{n_2\sqrt{2F_0+1}-n_1\sqrt{2-2F_0}}{\sqrt{(n_1+n_2)\sqrt{2F_0+1}\sqrt{2-2F_0}}})$.

We can apply a similar testing for $F_0 > F_1$. It is sufficient to replace the time allocation to $t_1 = 0$ $t_2 = \frac{t\sqrt{6}}{\sqrt{6}+\sqrt{2(1-F_0)}}$, $t_3 = \frac{t\sqrt{2(1-F_0)}}{\sqrt{6}+\sqrt{2(1-F_0)}}$. In this case, the likelihood ratio test with the risk probability α is almost equal to the test with the rejection region: $\{(n_2,n_3)|\frac{n_2}{n_2+n_3} \leq \frac{\sqrt{1-F_0}}{\sqrt{3}+\sqrt{1-F_0}} + \frac{\Phi^{-1}(\alpha)}{\sqrt{3}+\sqrt{1-F_0}}\sqrt{\frac{\sqrt{1-F_0}\sqrt{3}}{n_2+n_3}}\}$ concerning the null hypothesis $H_0 : F \leq F_0$. The p-value of this kind of tests is $\Phi(\frac{n_2\sqrt{3}-n_3\sqrt{1-F_0}}{\sqrt{(n_2+n_3)\sqrt{1-F_0}\sqrt{3}}})$.

Next, we consider the case where the dark count parameter δ is known but is not negligible, the Fisher information matrix is given by

$$\begin{pmatrix} \lambda(\frac{2\lambda t_1}{3(\lambda(2F+1)+6\delta)} + \frac{2\lambda t_2}{3(\lambda(2-2F)+6\delta)}) & \frac{\lambda(2F+1)}{3(\lambda(2F+1)+6\delta)}t_1 - \frac{\lambda(2-2F)}{3(\lambda(2-2F)+6\delta)}t_2 \\ \frac{\lambda(2F+1)}{3(\lambda(2F+1)+6\delta)}t_1 - \frac{\lambda(2-2F)}{3(\lambda(2-2F)+6\delta)}t_2 & \frac{2F+1}{\lambda(2F+1)+6\delta}\frac{2F+1}{6}t_1 + \frac{2-2F}{\lambda(2-2F)+6\delta}\frac{2-2F}{6}t_2 + \frac{1}{\lambda}t_3 \end{pmatrix}. \qquad (65)$$

Hence, from (46), the inverse of the minimum variance is equal to

$$f(t_1,t_2,t_3)$$
$$:=\lambda(\frac{2\lambda t_1}{3(\lambda(2F+1)+6\delta)} + \frac{2\lambda t_2}{3(\lambda(2-2F)+6\delta)}$$
$$- \frac{(\frac{\lambda(2F+1)}{3(\lambda(2F+1)+6\delta)}t_1 - \frac{\lambda(2-2F)}{3(\lambda(2-2F)+6\delta)}t_2)^2}{\frac{\lambda(2F+1)}{\lambda(2F+1)+6\delta}\frac{2F+1}{6}t_1 + \frac{\lambda(2-2F)}{\lambda(2-2F)+6\delta}\frac{2-2F}{6}t_2 + t_3}).$$

Then, we apply Lemmas 1 and 2 in Appendix A to $\frac{f(t_1,t_2,t_3)}{\lambda}$ with $a = \frac{2\lambda}{3(\lambda(2F+1)+6\delta)}$, $b = \frac{2\lambda}{3(\lambda(2-2F)+6\delta)}$,

$c = \frac{\lambda(2F+1)}{\lambda(2F+1)+6\delta}\frac{2F+1}{6}$, $d = \frac{\lambda(2-2F)}{\lambda(2-2F)+6\delta}\frac{2-2F}{6}$, and obtain the optimized value:

(i) coincidence and total flux

$$\max_{t_1+t_3=t} f(t_1,0,t_3) = \frac{4\lambda t}{((2F+1)+\sqrt{\frac{6(\lambda(2F+1)+6\delta)}{\lambda}})^2} \qquad (66)$$

(ii) anti-coincidence and total flux

and

$$\max_{t_2+t_3=t} f(0,t_2,t_3) = \frac{4\lambda t}{((2-2F)+\sqrt{\frac{6(\lambda(2-2F)+6\delta)}{\lambda}})^2} \tag{67}$$

(iii) coincidence and anti-coincidence

$$\max_{t_1+t_2=t} f(t_1,t_2,0) = \lambda t \left(\frac{\frac{\lambda(2F+1)}{3\sqrt{(\lambda(2F+1)+6\delta)(\lambda(2-2F)+6\delta)}} + \frac{\lambda(2-2F)}{3\sqrt{(\lambda(2F+1)+6\delta)(\lambda(2-2F)+6\delta)}}}{(2F+1)\sqrt{\frac{\lambda}{6(\lambda(2F+1)+6\delta)}} + (2-2F)\sqrt{\frac{\lambda}{6(\lambda(2-2F)+6\delta)}}} \right)^2$$

$$= \frac{2\lambda^2 t}{3(\lambda(2F+1)+6\delta)(\lambda(2-2F)+6\delta)} \left(\frac{3}{\frac{2F+1}{\sqrt{\lambda(2F+1)+6\delta}} + \frac{2-2F}{\sqrt{\lambda(2-2F)+6\delta}}} \right)^2$$

$$= \frac{6\lambda^2 t}{\left((2F+1)\sqrt{\lambda(2-2F)+6\delta} + (2-2F)\sqrt{\lambda(2F+1)+6\delta}\right)^2}. \tag{68}$$

The ratio of (66) to (68) is

$$\frac{3\lambda\left((2F+1)+\sqrt{\frac{6(\lambda(2F+1)+6\delta)}{\lambda}}\right)^2}{2\left((2F+1)\sqrt{\lambda(2-2F)+6\delta} + (2-2F)\sqrt{\lambda(2F+1)+6\delta}\right)^2}$$

$$= \frac{3}{2} \left(\frac{(2F+1)\sqrt{\lambda} + \sqrt{6(\lambda(2F+1)+6\delta)}}{(2F+1)\sqrt{\lambda(2-2F)+6\delta} + (2-2F)\sqrt{\lambda(2F+1)+6\delta}} \right)^2 > 1, \tag{69}$$

where the final inequality is derived in Appendix B. Therefore, the measurement using the coincidence and the anti-coincidence provides better test than that using the coincidence and the total flux, as in the case of $\delta = 0$.

Define δ_1 and the critical point $F_{\delta'}$ for the normalized dark count $\delta' = 6\delta/\lambda < \delta_1$ as

$$\sqrt{\delta_1 + 3} - \sqrt{\delta_1} = \sqrt{3/2}$$
$$\sqrt{1 + 2F_{\delta'} + \delta'} - \sqrt{2 - 2F_{\delta'} + \delta'} = \sqrt{3/2}.$$

The parameter δ_1 is calculated to be 0.375. As shown in Appendix C, the measurement using the coincidence and the anti-coincidence provides better test than that using the anti-coincidence and the total flux, if the fidelity is

smaller than the critical point $F_{\delta'}$:

$$\max_{t_1+t_2+t_3=t} f(t_1,t_2,t_3)$$

$$= \begin{cases} \frac{4\lambda^2 t}{((2-2F)\sqrt{\lambda}+\sqrt{6(\lambda(2-2F)+6\delta)})^2} & \text{if } F > F_{\delta'} \\ \frac{6\lambda^2 t}{\left((2F+1)\sqrt{\lambda(2-2F)+6\delta}+(2-2F)\sqrt{\lambda(2F+1)+6\delta}\right)^2} & \text{otherwise.} \end{cases} \tag{70}$$

The optimal time allocation is given by $t_1 = 0$, $t_2 = \frac{t\sqrt{6(\lambda(2-2F)+6\delta)}}{\sqrt{6(\lambda(2-2F)+6\delta)}+(2-2F)\sqrt{\lambda}}$, and $t_3 = \frac{t(2-2F)\sqrt{\lambda}}{\sqrt{6(\lambda(2-2F)+6\delta)}+(2-2F)\sqrt{\lambda}}$ for $F > F_{\delta'}$, and $t_1 = \frac{t(2-2F)\sqrt{\lambda(2F+1)+6\delta}}{(2-2F)\sqrt{\lambda(2F+1)+6\delta}+(2F+1)\sqrt{\lambda(2-2F)+6\delta}}$, $t_2 = \frac{t(2F+1)\sqrt{\lambda(2-2F)+6\delta}}{(2-2F)\sqrt{\lambda(2F+1)+6\delta}+(2F+1)\sqrt{\lambda(2-2F)+6\delta}}$, $t_3 = 0$ for $F \leq F_{\delta'}$. The critical point $F_{\delta'}$ for optimal time allocation increases with the normalized dark count as illustrated in Fig. 3.
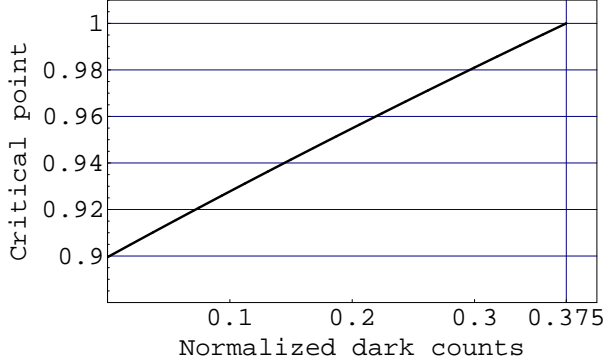
FIG. 3: The critical point $F_{\delta'}$ for optimal time allocation as a function of normalized dark counts δ'.

VIII. DESIGN II (λ: KNOWN, ONE STAGE)

In this section, we consider the case where λ is known. Then, the Fisher information is

$$\lambda\left(\frac{2\lambda t_1}{3(\lambda(2F+1)+6\delta)} + \frac{2\lambda t_2}{3(\lambda(2-2F)+6\delta)}\right). \quad (71)$$

The maximum value is calculated as

$$\max_{t_1+t_2+t_3=t}(71) = \begin{cases} \frac{2\lambda^2 t}{3(\lambda(2F+1)+6\delta)} & \text{if } F < \frac{1}{4} \\ \frac{2\lambda^2 t}{3(\lambda(2-2F)+6\delta)} & \text{if } F \geq \frac{1}{4}. \end{cases} \quad (72)$$

The above optimization shows that when $F \geq \frac{1}{4}$, the count on anti-coincidence ($t_1 = 0; t_2 = t; t_3 = 0$) is better than the count on coincidence ($t_1 = t; t_2 = 0; t_3 = 0$). In fact, Barbieri et $al.$[7] measured the sum of the counts on the anti-coincidence vectors $|HV\rangle, |VH\rangle, |DX\rangle, |XD\rangle, |RR\rangle, |LL\rangle$ to realize the entanglement witness in their experiment. In this case, the variance is $\frac{3(\lambda(2-2F)+6\delta)}{2\lambda^2 t}$. When we observe the sum of counts on anti-coincidence n_2, the estimated value of F is given by $1 + 3(\delta - \frac{n_2}{\lambda t})$, which is the solution of $(\lambda\frac{2-2F}{6} + \delta)t = n_2$. The likelihood ratio test with the risk probability α can be approximated by the test with the rejection region: $\{n_2|n_2 \leq (\frac{\lambda(1-F_0)}{3} + \delta)t + \Phi^{-1}(\alpha)\sqrt{(\frac{\lambda(1-F_0)}{3} + \delta)t}\}$ concerning the null hypothesis $H_0 : F \leq F_0$, which is also the UMP test. The p-value of likelihood ratio tests is $\Phi(\frac{n_2 - (\frac{\lambda(1-F_0)}{3}+\delta)t}{\sqrt{(\frac{\lambda(1-F_0)}{3}+\delta)t}})$.

When $F < \frac{1}{4}$, the optimal time allocation is $t_1 = t$, $t_2 = t_3 = 0$. The fidelity is estimated by $\frac{3n_2}{\lambda t} - \frac{1}{2}$. Its variance is $\frac{3(\lambda(2F+1)+6\delta)}{2\lambda^2 t}$. The likelihood ratio test with the risk probability α of the Poisson distribution

is almost equal to the test with the rejection region: $\{n_1|n_1 \geq (\lambda\frac{1+2F_0}{6} + \delta)t + \Phi^{-1}(1-\alpha)\sqrt{(\lambda\frac{1+2F_0}{6} + \delta)t}\}$ concerning the null hypothesis $H_0 : F \leq F_0$, which is also the UMP test. The p-value of likelihood ratio tests is $\Phi(\frac{-n_1 + (\lambda\frac{1+2F_0}{6}+\delta)t}{\sqrt{(\lambda\frac{1+2F_0}{6}+\delta)t}})$.

IX. COMPARISON OF THE ASYMPTOTIC VARIANCES

We compare the asymptotic variances of the following designs for time allocation, when the dark count δ parameter is zero.

(i) Modified visibility: The asymptotic variance is $\frac{(2F+1)(2-2F)}{\lambda t}$.

(iia) Design I (λ unknown). optimal time allocation between the counts on anti-coincidence and coincidence: The asymptotic variance is $\frac{(2F+1)(2-2F)(\sqrt{2F+1}+\sqrt{2-2F})^2}{6\lambda t}$.

(iib) Design I (λ unknown), optimal time allocation between the counts on anti-coincidence and the total flux: The asymptotic variance is $\frac{(2-2F)(\sqrt{6}+\sqrt{2-2F})^2}{4\lambda t}$.

(iiia) Design II (λ known), estimation from the count on anti-coincidence: The asymptotic variance is $\frac{3(2-2F)}{2\lambda t}$.

(iiib) Design II (λ known), estimation from the count on coincidence: The asymptotic variance is $\frac{3(2F+1)}{2\lambda t}$.

Fig. 4 shows the comparison, where the asymptotic variances in (iia)-(iiib) are normalized by the one in (i). The anti-coincidence measurement provides the best estimation for high ($F > 0.25$) fidelity. When λ is unknown, the measurement with the counts on anti-coincidence and the coincidence is better than that with the counts anti-coincidence and the total flux for $F < 0.899519$. For higher fidelity, the counts on anti-coincidence and total flux turns to be better, but the difference is small.

X. DESIGN III (λ: KNOWN, TWO STAGE)

A. Optimal Allocation

The comparison in the previous section shows that the measurement on the anti-coincidence vectors yields a better variance than the measurement on the coincidence vectors, when the fidelity is greater than 1/4 and the parameters λ and δ are known. We will explore further improvement in the measurement on the anti-coincidence vectors. In the previous sections, we allocate an equal time to the measurement on each of the anti-coincidence
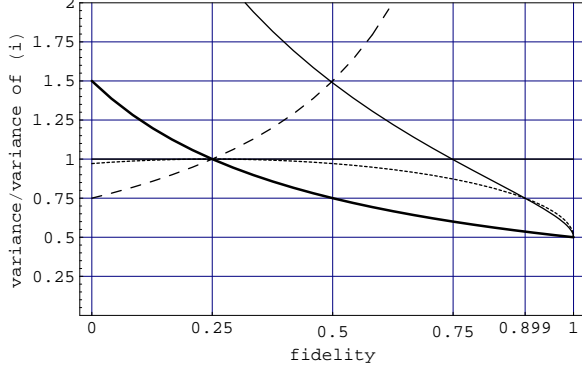
FIG. 4: Comparison of the designs for time allocation. The asymptotic variances normalized by the value of modified visibility method are shown as a function of fidelity, where dots: (iia), solid: (iib), thick: (iiia), and dash: (iiib).

vectors. Here we minimize the variance by optimizing the time allocation t_{HV}, t_{VH}, t_{DX}, t_{XD}, t_{RR}, and t_{LL} between the anti-coincidence vectors $B = \{|HV\rangle, |VH\rangle, |DX\rangle, |XD\rangle, |RR\rangle$, and $|LL\rangle\}$, under the restriction of the total measurement time: $\sum_{(x,y)\in B} t_{x,y} = t$. The number of the counts n_{xy} obeys Poisson distribution $\mathrm{Poi}((\lambda\mu_{xy} + \delta)t_{xy})$ with unknown parameter μ_{xy}. Then, the Fisher information matrix is the diagonal matrix with the diagonal elements $\{\frac{\lambda^2 t_{x,y}}{\lambda\mu_{x,y}+\delta}\}_{(x,y)\in B}$ Since we are interested in the parameter $1 - F = \frac{1}{2}(\sum_{(x,y)\in B}\mu_{x,y})$, the variance is given by

$$\frac{1}{4}\Big(\sum_{(x,y)\in B} \frac{\lambda\mu_{x,y} + \delta}{\lambda^2 t_{x,y}} \Big), \tag{73}$$

as mentioned in section V A. Under the restriction of the total measurement time, the minimum value of (73) is

$$\frac{(\sum_{(x,y)\in B} \sqrt{\lambda\mu_{x,y} + \delta})^2}{4\lambda^2 t}, \tag{74}$$

which is attained by the optimal time allocation

$$t_{xy} = \frac{(\lambda\sqrt{\mu_{xy}} + \delta)t}{\sum_{(x',y')\in B} \sqrt{\lambda\mu_{x',y'} + \delta}}, \tag{75}$$

which is called Neyman allocation and is used in sampling design[13]. The variance with the equal allocation is

$$\frac{3(\lambda(2 - 2F) + 6\delta)}{2\lambda^2 t} = \frac{3(\lambda(\sum_{(x,y)\in B}\mu_{x,y}) + 6\delta)}{2\lambda^2 t}. \tag{76}$$

The inequality (74) $\leq$ (76) can be derived from Schwartz's inequality of the vectors $(1,\ldots,1)$ and

$(\sqrt{\lambda\mu_{HV} + \delta},\ldots,\sqrt{\lambda\mu_{LL} + \delta})$. The equality holds if and only if $\mu_{HV} = \mu_{VH} = \mu_{DX} = \mu_{XD} = \mu_{RR} = \mu_{LL}$. Therefore, the Neyman allocation has an advantage over the equal allocation, when there is a bias in the parameters $\mu_{HV}, \mu_{VH}, \mu_{DX}, \mu_{XD}, \mu_{RR}, \mu_{LL}$. In other words, the Neyman allocation is effective when the expectation values of the counts on some vectors are larger than those on other vectors.

B. Two-stage Method

The optimal time allocation derived above is not applicable in the experiment, because it depends on the unknown parameters μ_{HV}, μ_{VH}, μ_{DX}, μ_{XD}, μ_{RR}, and μ_{LL}. In order to resolve this problem, we introduce a two-stage method, where the total measurement time t is divided into t_f for the first stage and t_s for the second stage under the condition of $t = t_f + t_s$. In the first stage, we measure the counts on each vectors for $t_f/6$ and estimate the expectation value for Neyman allocation on measurement time t_s. In the second stage, we measure the counts on a vector $|x_A y_B\rangle$ according to the estimated Neyman allocation. The two-stage method is formulated as follows.

(i) The measurement time for each vector in the first stage is given by $t_f/6$

(ii) In the second stage, we measure the counts on a vector $|x_A y_B\rangle$ with the measurement time $\tilde{t}_{xy}$ defined as

$$\tilde{t}_{xy} = \frac{m_{xy}}{\sum_{(x,y)\in B} \sqrt{m_{xy}}}(t - t_f)$$

where m_{xy} is the observed count in the first stage.

(iii) Define $\hat{\mu}_{xy}$ and $\hat{F}$ as

$$\hat{\mu}_{xy} = \frac{n_{xy}}{\lambda\tilde{t}_{xy}}, \quad \hat{F} = 1 - \frac{1}{2} \sum_{(x,y)\in B} \hat{\mu}_{x,y},$$

where $n_{x,y}$ is the number of the counts on $|x_A y_B\rangle$ for $\tilde{t}_{xy}$. Then, we can estimate the fidelity by $\hat{F}$.

(iv) Finally, we apply the test $\phi_{\mathrm{mod},\alpha}$ given in Section IV D to the two hypotheses given as

$$H_0 : \vec{w} \cdot \vec{\mu} \geq c_0 \text{ versus } H_1 : \vec{w} \cdot \vec{\mu} < c_0, \tag{77}$$

where $w_{x,y} := \frac{1}{2\lambda\tilde{t}_{x,y}}$ and $c_0 := 1 - F_0$.

XI. CONCLUSION

We have formulated the hypothesis testing scheme to test the entanglement in the Poisson distribution framework. Our statistical method can handle the fluctuation in the experimental data more properly in a realistic setting. It has been shown that the optimal time allocation improves the test: the measurement time should be allocated preferably to the anti-coincidence vectors. This

test is valid even if the dark count exists. This design is particularly useful for the experimental test, because the optimal time allocation depends only on the threshold of the test. We don't need any further information of the probability distribution and the tested state. The test can be further improved by optimizing time allocation between the anti-coincidence vectors, when the error from the maximally entangled state is anisotropic. However, this time allocation requires the expectation values on the counts on coincidence, so that we need to apply the two stage method.

APPENDIX A: OPTIMIZATION OF FISHER INFORMATION

In this section, we maximize the quantities appearing in Fisher information.

Lemma 1 *The equation*

$$\max_{t_1,t_3\geq 0,\ ct_1+t_3=t} at_2 - \frac{act_1^2}{ct_1+t_3} = \frac{at}{(\sqrt{c}+1)^2} \quad \text{(A1)}$$

holds and the maximum value is attained when $t_1 = \frac{t}{\sqrt{c}+1}$, $t_3 = \frac{\sqrt{c}t}{\sqrt{c}+1}$.

Proof: Letting $x := ct_1 + t_3$, we have $t_1 = \frac{x-t}{c-1}$. Then,

$$at_1 - \frac{act_1^2}{ct_1+t_3} = \frac{a}{(c-1)^2}\left(-x - \frac{ct^2}{x} + (c+1)t\right).$$

Hence, the maximum is attained at $x = \sqrt{c}t$, i.e., $t_1 = \frac{t}{\sqrt{c}+1}$ and $t_3 = \frac{t\sqrt{c}}{\sqrt{c}+1}$. Thus,

$$\max_{t_1,t_3\geq 0,\ ct_1+t_3=t} at_1 - \frac{act_1^2}{ct_1+t_3}$$
$$= \frac{a}{(c-1)^2}\left(-2\sqrt{c}t + (c+1)t\right) = \frac{at}{(\sqrt{c}+1)^2}.$$

∎

Lemma 2 *The equation*

$$\max_{t_1,t_2\geq 0,\ t_1+t_2=t} at_1 + bt_2 - \frac{(\sqrt{ac}t_1 - \sqrt{bc}t_2)^2}{ct_1+dt_2}$$
$$= \frac{t(\sqrt{ad}+\sqrt{bc})^2}{(\sqrt{c}+\sqrt{d})^2}. \quad \text{(A2)}$$

holds, and this maximum value is attained when $t_1 = \frac{t\sqrt{d}}{\sqrt{c}+\sqrt{d}}$, $t_2 = \frac{t\sqrt{c}}{\sqrt{c}+\sqrt{d}}$.

Proof: Letting $x := ct_1 + dt_2$, we have $t_1 = \frac{dt-x}{d-c}$ and $t_2 = \frac{x-ct}{d-c}$. Then,

$$at_1 + bt_2 - \frac{(\sqrt{ac}t_1 - \sqrt{bc}t_2)^2}{ct_1+dt_2}$$
$$= \left(\frac{\sqrt{ad}+\sqrt{bc}}{d-c}\right)^2\left((c+d)t - x - \frac{cdt^2}{x}\right).$$

Hence, the maximum is attained at $x = \sqrt{cd}t$, i.e., $t_1 = \frac{t\sqrt{d}}{\sqrt{c}+\sqrt{d}}$ and $t_2 = \frac{t\sqrt{c}}{\sqrt{c}+\sqrt{d}}$. Thus,

$$\max_{t_1,t_2\geq 0,\ t_1+t_2=t} at_1 + bt_2 - \frac{(\sqrt{ac}t_1 - \sqrt{bd}t_2)^2}{ct_1+dt_2}$$
$$= \left(\frac{\sqrt{ad}+\sqrt{bc}}{d-c}\right)^2\left((c+d)t - 2\sqrt{cd}t\right) = \frac{t(\sqrt{ad}+\sqrt{bc})^2}{(\sqrt{c}+\sqrt{d})^2}.$$

∎

Further, three-parameter case can be maximized as follows.

Lemma 3 *The maximum value*

$$\max_{t_1,t_2,t_3\geq 0,\ t_1+t_2+t_3=t} at_1 + bt_2 - \frac{(\sqrt{ac}t_1 - \sqrt{bd}t_2)^2}{ct_1+dt_2+t_3}$$

is equal to the maximum among three values

$$\max_{t_1,t_3\geq 0,\ ct_1+t_3=t} at_2 - \frac{act_1^2}{ct_1+t_3}, \quad \max_{t_2,t_3\geq 0,\ ct_2+t_3=t} at_2 -$$
$$\frac{bdt_2^2}{dt_2+t_3}, \quad \max_{t_1,t_2\geq 0,\ t_1+t_2=t} at_1 + bt_2 - \frac{(\sqrt{ac}t_1 - \sqrt{bd}t_2)^2}{ct_1+dt_2}.$$

Proof: Define two parameters $x := ct_1 + dt_2 + t_3$ and $y := \sqrt{cd}t_1 - \sqrt{bd}t_2$. Then, the range of x and y forms a convex set. Since

$$t_1 = \frac{\sqrt{bd}(x-t) + (d-1)y}{\sqrt{bd}(c-1) + \sqrt{ac}(d-1)},$$
$$t_2 = \frac{\sqrt{ac}(x-t) - (c-1)y}{\sqrt{ac}(c-1) + \sqrt{bd}(d-1)}.$$

Hence,

$$at_1 + bt_2 - \frac{(\sqrt{ac}t_1 - \sqrt{bd}t_2)^2}{ct_1+dt_2+t_3}$$
$$= \left(\frac{a\sqrt{bd}}{\sqrt{bd}(c-1)+\sqrt{ac}(d-1)}\right.$$
$$\left. + \frac{b\sqrt{ac}}{\sqrt{ac}(c-1)+\sqrt{bd}(d-1)}\right)(x-t)$$
$$+ \left(\frac{a(d-1)}{\sqrt{bd}(c-1)+\sqrt{ac}(d-1)}\right.$$
$$\left. - \frac{b(c-1)}{\sqrt{ac}(c-1)+\sqrt{bd}(d-1)}\right)y - \frac{y^2}{x}$$
$$= -\frac{1}{x}\left(y - \frac{1}{2}Bx\right)^2 + \left(\frac{B^2}{4}+A\right)x - At,$$

where $A := \frac{a\sqrt{bd}}{\sqrt{bd}(c-1)+\sqrt{ac}(d-1)} + \frac{b\sqrt{ac}}{\sqrt{ac}(c-1)+\sqrt{bd}(d-1)}$, $B :=$ $\frac{a(d-1)}{\sqrt{bd}(c-1)+\sqrt{ac}(d-1)} - \frac{b(c-1)}{\sqrt{ac}(c-1)+\sqrt{bd}(d-1)}$. Applying Lemma 4, we obtain this lemma. ∎

Lemma 4 *Define the function $f(x,y) := -\frac{1}{x}(y-\alpha x)^2 + \beta x$ on a closed convex set C. The maximum value is realized at the boundary bdC.*

Proof: The condition can be classified to two cases: i) $bdC \cap \{y = \alpha x\} = \emptyset$, ii) $bdC \cap \{y = \alpha x\} \neq \emptyset$. In the case i), when fix x is fixed, $\max_{y:(x,y)\in C} f(x,y) = \max_{y:(x,y)\in bdC} f(x,y)$. Then, we obtain $\max_{(x,y)\in C} f(x,y) = \max_{(x,y)\in bdC} f(x,y)$. In the case ii), when $(x,\alpha x) \in C$, $\max_{y:(x,y)\in C} f(x,y) = f(x,\alpha x) = \beta x$. Hence, $\max_{x:(x,\alpha x)\in C} \max_{y:(x,y)\in C} f(x,y) = \max_{x:(x,\alpha x)\in C} \beta x$ This maximum is attained at $x = \max_x\{x| : (x,\alpha x) \in C\}$ or $x = \min_x\{x| : (x,\alpha x) \in C\}$. These point belongs to the boundary bdC. Further, $\max_{x:(x,\alpha x)\notin C} \max_{y:(x,y)\in C} f(x,y) = \max_{x:(x,\alpha x)\in C} \max_{y:(x,y)\in bdC} f(x,y)$. Thus, the proof is completed. ∎

APPENDIX B: PROOF OF INEQUALITIES (61) AND (69)

It is sufficient to show

$$\sqrt{\frac{3}{2}}\left((2F+1)\sqrt{\lambda} + \sqrt{6(\lambda(2F+1)+6\delta)}\right)$$
$$- \left((2F+1)\sqrt{\lambda(2-2F)+6\delta}\right.$$
$$\left. + (2-2F)\sqrt{\lambda(2F+1)+6\delta}\right) > 0. \qquad (B1)$$

By putting $\delta' := \frac{6\delta}{\lambda}$, the LHS is evaluated as

$$\frac{\text{LHS of (B1)}}{\sqrt{\lambda}}$$
$$=\sqrt{\frac{3}{2}}(2F+1) + 3\sqrt{(2F+1)+\delta'}$$
$$- (2F+1)\sqrt{(2-2F)+\delta'} - (2-2F)\sqrt{(2F+1)+\delta'}$$
$$=\sqrt{\frac{3}{2}}(2F+1) + (2F+1)\sqrt{(2F+1)+\delta'}$$
$$- (2F+1)\sqrt{(2-2F)+\delta'}$$
$$=(2F+1)\left(\sqrt{\frac{3}{2}} + \sqrt{(2F+1)+\delta'} - \sqrt{(2-2F)+\delta'}\right).$$

Since $0 \leq F \leq 1$, we have

$$\sqrt{\frac{3}{2}} + \sqrt{(2F+1)+\delta'} - \sqrt{(2-2F)+\delta'}$$
$$\geq \sqrt{\frac{3}{2}} + \sqrt{1+\delta'} - \sqrt{2+\delta'}.$$

Further, the function $\delta' \to \sqrt{1+\delta'} - \sqrt{2+\delta'}$ ($\delta' \in [0,\infty]$) has the minimum $\sqrt{1} - \sqrt{2} > -1 > -\sqrt{\frac{3}{2}}$ at $\delta' = 0$. Hence, $\frac{\text{LHS of (B1)}}{\sqrt{\lambda}} > 0$.

APPENDIX C: PROOF OF EQUATIONS (62) AND (70)

It is sufficient to show that

$$\sqrt{\frac{3}{2}}\left((2-2F)\sqrt{\lambda} + \sqrt{6(\lambda(2-2F)+6\delta)}\right)$$
$$- \left((2F+1)\sqrt{\lambda(2-2F)+6\delta}\right.$$
$$\left. + (2-2F)\sqrt{\lambda(2F+1)+6\delta}\right) > 0 \qquad (C1)$$

if and only if $\frac{6\delta}{\lambda} < \delta_1$ and $F \geq F_{\frac{6\delta}{\lambda}}$. By putting $\delta' := \frac{6\delta}{\lambda}$, the LHS of (C1) is evaluated as

$$\frac{\text{LHS of (C1)}}{\sqrt{\lambda}}$$
$$=\sqrt{\frac{3}{2}}(2-2F) + 3\sqrt{(2-2F)+\delta'}$$
$$- (2F+1)\sqrt{(2-2F)+\delta'} - (2-2F)\sqrt{(2F+1)+\delta'}$$
$$=\sqrt{\frac{3}{2}}(2-2F) + (2-2F)\sqrt{(2-2F)+\delta'}$$
$$- (2-2F)\sqrt{(2F+1)+\delta'}$$
$$=(2-2F)\left(\sqrt{\frac{3}{2}} + \sqrt{(2-2F)+\delta'} - \sqrt{(2F+1)+\delta'}\right).$$

Since $0 \leq F \leq 1$ and $\delta \geq 0$,

$$\sqrt{\frac{3}{2}} + \sqrt{(2-2F)+\delta'} - \sqrt{(2F+1)+\delta'} > 0$$

if and only if $\delta_1 > \delta'$ and $F > F_{\delta'}$.

APPENDIX D: PROOF OF (38)

Define $\tilde{\mu}_i$ by

$$\min_{\vec{w}\cdot\vec{\mu}'=c_0} D(\text{Poi}(0,\ldots,0,\tilde{\mu}_i,0,\ldots,0)\|\text{Poi}(\vec{\mu}')) = R.$$

In fact, when $w_i a < c_0$,

$$\min_{\mu_i'\geq 0:\ \vec{w}\cdot\vec{\mu}'=c_0} D(\text{Poi}(0,\ldots,0,a,0,\ldots,0)\|\text{Poi}(\vec{\mu}'))$$
$$= \min_{\mu_i'\geq 0:\ \vec{w}\cdot\vec{\mu}'=c_0} \sum_{j=1}^{m} \mu_j' - a + a\log\frac{a}{\mu_i'}$$
$$= \min_{\alpha\geq 0,\beta\geq 0:\ \mu_i\alpha+\mu_M\beta=c_0} \alpha + \beta - a + a\log\frac{a}{\alpha}$$
$$= \begin{cases} \frac{c_0}{w_i} - a + a\log\frac{aw_i}{c_0} & \text{if } a \geq \frac{c_0(w_M-w_i)}{w_M w_i} \\ \frac{c_0}{w_M} + a\log\frac{w_M-w_i}{w_M} & \text{if } a < \frac{c_0(w_M-w_i)}{w_M w_i}. \end{cases}$$

This value is monotone decreasing concerning a. When $a = \frac{c_0(w_M - w_i)}{w_M w_i}$, this value is $\frac{c_0}{w_M} - \frac{c_0(w_M - w_i)}{w_i w_M} \log \frac{w_M - w_i}{w_M}$. Hence, the value $\tilde{\mu}_i$ coincides with the the value $\tilde{\mu}_i$ defined by (36) and (37).

Thus, the relation (38) follows from the relation

$$
\min_{\vec{w} \cdot \vec{\mu}' = c_0} D(\text{Poi}(p_1 a_1, \ldots, p_m a_m) \| \text{Poi}(\vec{\mu}'))
$$
$$
\leq \sum_{i=1}^{m} p_i \min_{\vec{w} \cdot \vec{\mu}' = c_0} D(\text{Poi}(0, \ldots, 0, a_i, 0, \ldots, 0) \| \text{Poi}(\vec{\mu}')).
$$

We choose $\vec{\mu}_i'$ such that $\min_{\vec{w} \cdot \vec{\mu}' = c_0} D(\text{Poi}(0, \ldots, 0, a_i, 0, \ldots, 0) \| \text{Poi}(\vec{\mu}')) = D(\text{Poi}(0, \ldots, 0, a_i, 0, \ldots, 0) \| \text{Poi}(\vec{\mu}_i'))$. Then, the above inequality follows from Lemma 5 in the following way:

$$
\sum_{i=1}^{m} p_i \min_{\vec{w} \cdot \vec{\mu}' = c_0} D(\text{Poi}(0, \ldots, 0, a_i, 0, \ldots, 0) \| \text{Poi}(\vec{\mu}'))
$$
$$
= \sum_{i=1}^{m} p_i D(\text{Poi}(0, \ldots, 0, a_i, 0, \ldots, 0) \| \text{Poi}(\vec{\mu}_i'))
$$
$$
\geq D(\text{Poi}(p_1 a_1, \ldots, p_m a_m) \| \text{Poi}(\sum_{i=1}^{m} p_i \vec{\mu}_i'))
$$
$$
\geq \min_{\vec{w} \cdot \vec{\mu}' = c_0} D(\text{Poi}(p_1 a_1, \ldots, p_m a_m) \| \text{Poi}(\vec{\mu}')).
$$

Lemma 5 *Any real number* $0 \leq p \leq 1$ *and any four sequence of positive numbers* (μ_i), (ν_i), (μ_i'), *and* (ν_i') *satisfy*

$$
p \left(\sum_{i=1}^{m} (\mu_i - \nu_i) + \sum_{i=1}^{m} \nu_i \log \frac{\nu_i}{\mu_i} \right)
$$
$$
+ (1-p) \left(\sum_{i=1}^{m} (\mu_i' - \nu_i') + \sum_{i=1}^{m} \nu_i' \log \frac{\nu_i'}{\mu_i'} \right)
$$
$$
\geq \sum_{i=1}^{m} ((p\mu_i + (1-p)\mu_i') - p\nu_i + (1-p)\nu_i')
$$
$$
+ \sum_{i=1}^{m} (p\nu_i + (1-p)\nu_i') \log \frac{(p\nu_i + (1-p)\nu_i')}{(p\mu_i + (1-p)\mu_i')}.
$$

Proof: It is sufficient to show

$$
p \left(\sum_{i=1}^{m} \nu_i \log \frac{\nu_i}{\mu_i} \right) + (1-p) \left(\sum_{i=1}^{m} \nu_i' \log \frac{\nu_i'}{\mu_i'} \right)
$$
$$
\geq \sum_{i=1}^{m} (p\nu_i + (1-p)\nu_i') \log \frac{(p\nu_i + (1-p)\nu_i')}{(p\mu_i + (1-p)\mu_i')}.
$$

The convexity of $-\log$ implies that

$$
- \log \left(\frac{(p\mu_i + (1-p)\mu_i')}{(p\nu_i + (1-p)\nu_i')} \right)
$$
$$
= - \log \left(\frac{p\nu_i}{(p\nu_i + (1-p)\nu_i')} \frac{\mu_i}{\nu_i} + \frac{(1-p)\nu_i'}{(p\nu_i + (1-p)\nu_i')} \frac{\mu_i'}{\nu_i'} \right)
$$
$$
\leq \frac{p\nu_i}{(p\nu_i + (1-p)\nu_i')} \cdot - \log \left(\frac{\mu_i}{\nu_i} \right)
$$
$$
+ \frac{(1-p)\nu_i'}{(p\nu_i + (1-p)\nu_i')} \cdot - \log \left(\frac{\mu_i'}{\nu_i'} \right).
$$

Hence,

$$
p \left(\sum_{i=1}^{m} \nu_i \log \frac{\nu_i}{\mu_i} \right) + (1-p) \left(\sum_{i=1}^{m} \nu_i' \log \frac{\nu_i'}{\mu_i'} \right)
$$
$$
= \sum_{i=1}^{m} (p\nu_i + (1-p)\nu_i') \left(\frac{p\nu_i}{(p\nu_i + (1-p)\nu_i')} \cdot - \log \left(\frac{\mu_i}{\nu_i} \right) \right.
$$
$$
+ \left. \frac{(1-p)\nu_i'}{(p\nu_i + (1-p)\nu_i')} \cdot - \log \left(\frac{\mu_i'}{\nu_i'} \right) \right)
$$
$$
\geq - \sum_{i=1}^{m} (p\nu_i + (1-p)\nu_i') \log \frac{(p\nu_i + (1-p)\nu_i')}{(p\mu_i + (1-p)\mu_i')}.
$$

∎

APPENDIX E: PROOF OF (48)

Considering the shape of the graph $\frac{x}{\sqrt{y}} = a$, we can show that the minimum value $\min_{(x,y) \in C} \frac{x}{\sqrt{y}}$ can be attained by the boundary of C. Hence the boundary of the convex set $Co(R)$ is included by the union $\cup_{i \neq j} l_{i,j}$ of the lines $l_{i,j} = \{(t x_i(R) + (1-t) x_j(R), t y_i(R) + (1-t) y_j(R)) | 0 \leq t \leq 1\}$. Taking the derivative of $\frac{t x_i(R) + (1-t) x_j(R)}{\sqrt{t y_i(R) + (1-t) y_j(R)}}$ concerning t, we obtain

$$
\min_{t \in [0,1]} \frac{t x_i(R) + (1-t) x_j(R)}{\sqrt{t y_i(R) + (1-t) y_j(R)}} = z_{i,j}(R). \qquad (E1)
$$

Hence, we obtain (48).

[1] C.H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, *Phys. Rev. Lett.* **70**, 1895 (1993).

[2] H.-J. Briegel, W. Dur, J.I. Cirac, and P. Zoller, *Phys.*

Rev. Lett., **81**, 5932 (1998).

[3] C. W. Helstrom, *Quantum detection and estimation theory*, Academic Press (1976).

[4] A. S. Holevo, *Probabilistic and statistical aspects of quan-*

tum theory, North-Holland Publishing (1982).

[5] M. Hayashi, *Asymptotic Theory of Quantum Statistical Inference: Selected Papers*, World Scientific (2005).

[6] A. G. White, D. F. V. James, P. H. Eberhard, and P. G. Kwiat, *Phys. Rev. Lett.*, **83**, 3103 (1999).

[7] M. Barbieri, F. De Martini, G. Di Nepi, P. Mataloni, G. M. D'Ariano, and C. Macchiavello, *Phys. Rev. Lett.*, **91**, 227901 (2003).

[8] Y. Tsuda, K. Matsumoto, and M. Hayashi. "Hypothesis testing for a maximally entangled state," quant-ph/0504203.

[9] E. L. Lehmann, *Testing statistical hypotheses*, Second edition. Wiley (1986).

[10] P. G. Kwiat, E. Waks, A. G. White, I. Appelbaum, and P.H. Eberhard, *Phys. Rev. A*, **60**, 773(R) (1999).

[11] K. Usami, Y. Nambu, Y. Tsuda, K. Matsumoto, and K. Nakamura, "Accuracy of quantum-state estimation utilizing Akaike's information criterion," *Phys. Rev. A*, **68**, 022314 (2003).

[12] S. Amari and H. Nagaoka, *Methods of Information Geometry,* (AMS & Oxford University Press, 2000).

[13] W. G. Cochran, *Sampling Techniques*, third edition, John Wiley, (1977).

[14] M. Hayashi, B.-S. Shi, A. Tomita, K. Matsumoto, Y. Tsuda, and Y.-K. Jiang: "Hypothesis testing for an entangled state produced by spontaneous parametric down conversion," quant-ph/0603254.

[15]